

MPEG-7 – Eine Übersicht

Seminararbeit

vorgelegt am

Lehrstuhl für Praktische Informatik IV

Prof. Dr. W. Effelsberg

Universität Mannheim

im

Juni 2006

von

Victor Schan

aus Ketsch

Betreuer

Stephan Kopf

Inhaltsverzeichnis

1. Einleitung	1
2. MPEG-7 - Überblick	2
2.1. Aufbau des Standards	3
2.2. Deskriptoren	5
2.3. Description Schemes	6
2.4. Media Description Schemes	7
3. Ausgewählte Deskriptoren	8
3.1. Color Structure Descriptor	8
3.2. Dominant Color Descriptor	9
3.3. Homogeneous Texture Descriptor	10
3.4. Contour Shape Descriptor	11
3.5. Distanzmaß	12
4. Fazit	13
5. Literaturverzeichnis	i

Abbildungen:

Abbildung 1: MPEG-7 Umfang [Höynck, 2004].....	2
Abbildung 2: Description Scheme [Hönyck, 2004].....	5
Abbildung 3: Media Desription Schemes [Martinez, 2004]	6
Abbildung 4: Color Structure Descriptor.....	8
Abbildung 5: Dominant Color Structure.....	9
Abbildung 6: Homogeneous Texture Descriptor	10
Abbildung 7: Kontur eines gehenden Menschen [Kopf]	10
Abbildung 8: Krümmungsbasierter Skalenraum [Kopf].....	11

1. Einleitung

Audiovisuelle Informationen spielen in unserer Gesellschaft eine große Rolle und werden seit einigen Jahren digital aufgezeichnet und gespeichert. Desweiteren hat man damit begonnen, analoge Multimediadaten, welche sich über Jahrzehnte hinweg in den Archiven der Rundfunk- und Radioanstalten angesammelt haben, zu digitalisieren und auf neuen Medien zu speichern: Tendenz steigend. Hinzu kommen täglich Unmengen von Einzelbilder, 3D-Modellen, Musik und Sprache.

Um mit diesen stetig wachsenden Informationsmassen umzugehen, benötigt man Werkzeuge, welche diese Daten interpretieren und einheitlich speichern können. Man muss gezielt nach Inhaltsmerkmalen suchen können, um sich in den großen Multimediabibliotheken überhaupt noch zurecht finden zu können, was durch die automatische Analyse von audiovisuellen Signalen möglich wird.

Im Oktober 1996 begannen die Mitglieder der Motion Picture Experts Group (MPEG), die Interessengruppen wie Rundfunkanstalten, Elektronikhersteller, Verlage und Telekommunikationsanbieter repräsentieren, einen neuen Standard dafür zu entwickeln.

2. MPEG-7 - Überblick

Das „Multimedia Content Description Interface“ (kurz MPEG-7) ist im Vergleich zu seinen Vorgängern MPEG-1 und MPEG-2 kein Kompressionsstandard, sondern eine einheitliche Sprache zur Beschreibung von multimedialen Daten, deren Final Draft als International Standard im September 2001 verabschiedet wurde. Dieser neue Standard definiert eine Schnittstelle zwischen Merkmalsextraktion und Anwendungen, und stellt damit ein Format bereit, welches externe Programme zur Speicherung der extrahierten Merkmale verwenden können [Richter, 2003].

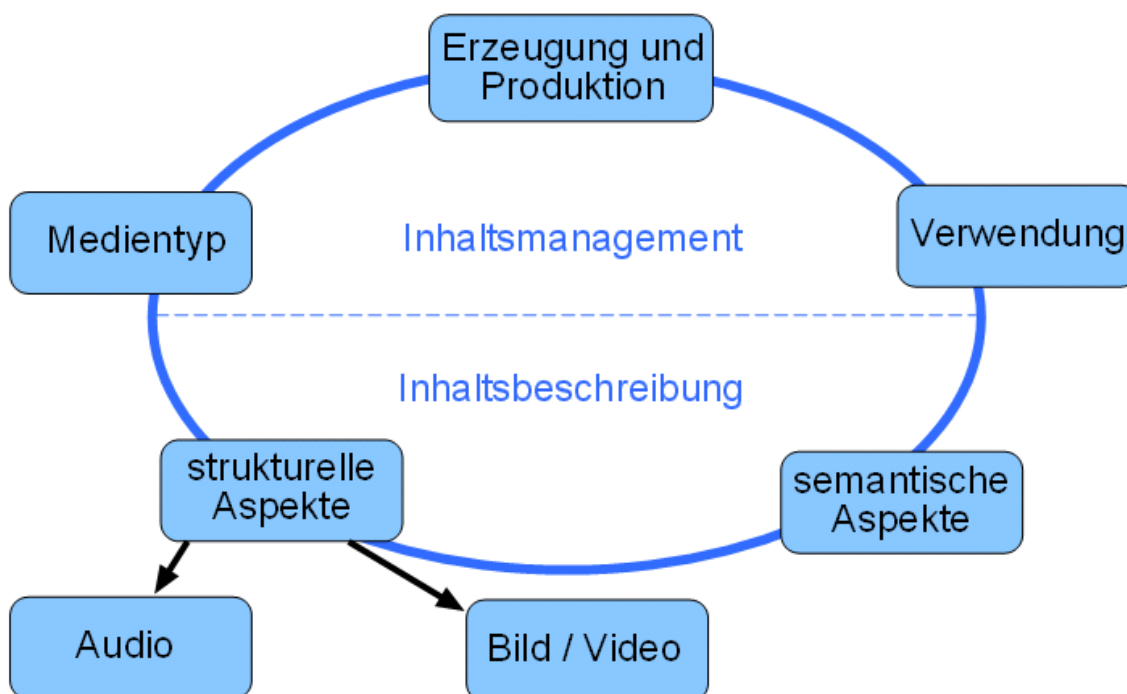


Abbildung 1: MPEG-7 Umfang [Höyneck, 2004]

MPEG-7 bietet umfangreiche Beschreibungsstrukturen, welche aus Basisbausteinen bestehen, die zu Schemas zusammengefasst werden und so den eigenen Ansprüchen angepasst werden können. Damit können dann nicht nur die reinen AV-Merkmale wie Farbe, Textur oder Frequenz beschrieben werden, sondern es können alle anderen nötigen Metadaten erfasst werden. Abbildung 1 zeigt die Wichtigsten: z.B. sind bei der Erzeugung und Produktion von Dokumentationen der Ort, die Zeit, anwesende Personen oder auch der Produktionsleiter wichtige Metadaten. Wenn genau festgelegt werden muss, was mit dem Material später passieren soll und ob es eventuelle rechtliche Einschränkungen bei der Verwertung gibt, fällt dies auch unter

Inhaltsmanagement. Der Inhaltsbeschreibung sind ebenfalls keine Grenzen gesetzt, sei es nun bei strukturellen Aspekten der Multimediadaten und Low-Level-Informationen wie z.B. Farben und Frequenzen oder bei semantischen Zusammenhängen, z.B. zwischen einzelnen Szenen und Objekten.

MPEG-7 ist auf kein Medium und kein Format beschränkt und auch für zukünftige Formate offen. Von Bildern, Grafiken und 3D-Modelle über Audio, Sprache und Video kann alles beschrieben werden, sogar wie man einzelne Elemente in einer Multimediapräsentation miteinander verknüpfen kann und welche semantischen Verbindungen zu anderen Multimediadokumenten bestehen.

2.1. Aufbau des Standards

Der MPEG-7 Standard ist in 10 Teilen organisiert:

1. *Systems* beschreibt Formate, in denen MPEG-7-Dokumente gespeichert und über ein Kommunikationsnetzwerk transportiert werden können. Ein wichtiger Bestandteil von *Systems* ist der BiM-Codec (Binary Format MPEG-7 encoder und decoder), der es ermöglicht, MPEG-7-Dokumente zu komprimieren. Das Dokument wird dabei in Fragmente zerlegt, die in einzelne Pakete („Access Units“) gruppiert werden. MPEG-7-Ströme werden getrennt oder synchronisiert mit denen von ihnen beschriebenen Medienströmen (z.B. MPEG-2/4) übertragen.
2. *Description Definition Language (DDL)* beschreibt die Syntax der MPEG-7-Beschreibungsstruktur und baut auf die von W3C standardisierte XML-Schema-Sprache auf. Zusätzlich zu dieser wurden hauptsächlich Datentypen zur Beschreibung von Zeitmarken und Matrizendimensionalitäten definiert, die in audiovisuellen Deskriptoren genutzt werden. Die Verwendung der DDL zur Definition der MPEG-7-Deskriptoren und Description Schemes ermöglicht es dem Anwender, basierend auf dem Standard, eigene applikationsspezifische Erweiterungen zu definieren.
3. Grundlegende visuelle Deskriptoren decken vor allem den automatisch erfassbaren Teil der Beschreibungen für Merkmale der Bilder und Videos ab. Typische visuelle Deskriptoren sind die Formdeskriptoren von Regionen,

Deskriptoren, die dominante Farben in Bildern oder Bildregionen spezifizieren, Deskriptoren für lokale oder globale Farbverteilungen, für Texturen und für Bewegungsbeschreibungen in Bildsequenzen.

4. Grundlegende *Audiodeskriptoren* spezifizieren beispielsweise die Merkmale von Audiosignalen, die Sprache oder Musik enthalten. Dazu gehören beispielsweise Beschreibungen von Merkmalen der Melodie oder Phonemnetzen. Es können aber auch die Eigenschaften der Erzeugerinstrumente beschrieben werden, wie z.B. die Klangfarbe einer Violine.
5. Die *Multimedia Description Schemes* spezifizieren eine Bibliothek an Beschreibungsstrukturen, aus denen Teile für eine Beschreibung instantiiert werden können. Die generisch aufgebauten Beschreibungsschemata ermöglichen semantische Beschreibungen von Medieninhalten und bilden einen Beschreibungsrahmen, in den die audio-visuellen Deskriptoren eingefügt werden können.
6. Die *Referenzsoftware* enthält Beispielimplementierungen für die Erzeugung und Verarbeitung der MPEG-7-Deskriptoren und Description Schemes. Die Software wurde im Standardisierungsprozess zur Validierung der standardisierten Technologien eingesetzt. Die Referenzsoftware ist unter www.lis.e-technik.tu-muenchen.de/research/bv/topics/mmdb/e_mpeg7.html verfügbar.
7. *Konformität* definiert Vorschriften für Konformitätstests der MPEG-7-Beschreibungen.
8. *Extraktion und Benutzung* stellt informative Beispiele für Instantiierung von Beschreibungsschemata und Ansätze für die Extraktion von Beschreibungen aus Medienströmen zur Verfügung. Zudem wurden auch Beispiele für den Gebrauch von Beschreibungen entworfen [Kosch, 2003].
9. *Profiles and Levels* sammelt Standardprofile und –komplexitätsstufen aus den vorangegangenen Teilen, wobei man sich hierbei auf die Description Definition Language, Visuelle Deskriptoren, Audio Deskriptoren und Multimedia Description Schemes konzentriert. Mit selbsterstellten Profilen kann man den Umfang und die Komplexität seinen eigenen Präferenzen entsprechend anpassen.

10. In *Schema Definition* werden die kompletten MPEG-7 Schematas zusammengefasst, die aus verschiedenen Standards, Corrigenden und Nachbesserungen stammen [Martinez, 2004].

2.2. Deskriptoren

Deskriptoren sind die kleinsten Bausteine in einem MPEG-7 Dokument. Sie beschreiben in kompakter Form einzelne Eigenschaften von multimedialen Daten. Die ursprünglichen Informationen können aus einem Deskriptor nicht wiedergewonnen werden, da sie diese nur beschreiben und dadurch eine, im Auge des Betrachters liegende, bestimmte Bedeutung wiedergeben.

Eine Instanz eines Deskriptors heißt Description Value (DV) und ist mit einem entsprechenden Wert belegt, der eine Ausprägung des zu beschreibenden Merkmals darstellt, ähnlich den Datentypen in der Programmierung. Ein Beispiel für einen Deskriptor wäre z.B. die Grundfarbe einer einfarbigen Fläche eines Bildes.

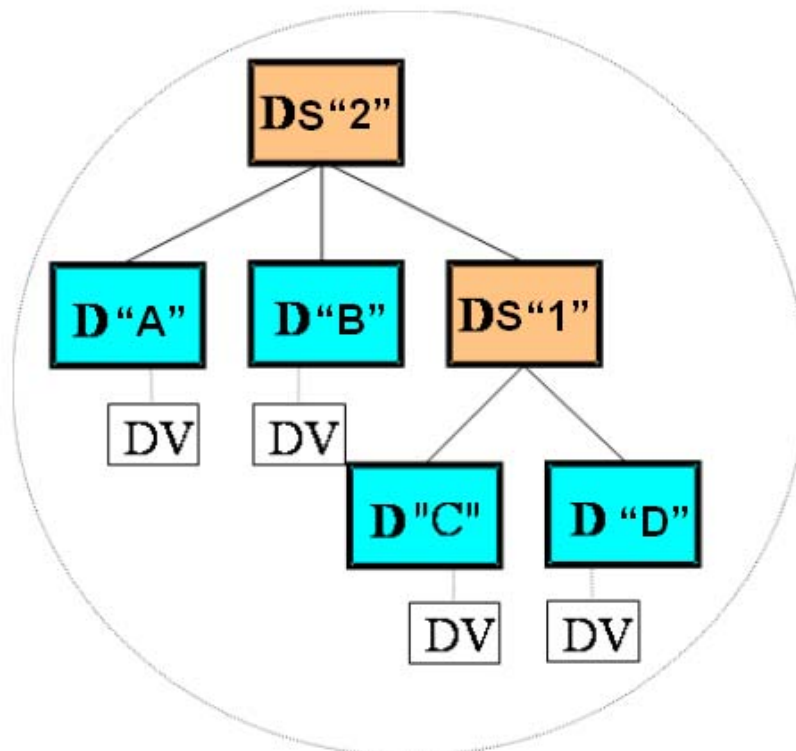


Abbildung 2: Description Scheme [Hönyck, 2004]

2.3. Description Schemes

Ein Description Scheme spezifiziert die Struktur und die Semantik zwischen verschiedenen Deskriptoren und anderen Description Schemes. Während Deskriptoren für einfache simple Audio-und Videoeigenschaften eingesetzt werden, wurden MPEG-7 Description Schemes hingegen für komplexe Eigenschaften wie Regionen, Segmente, Objekte, Ereignisse und andere Metadaten geschaffen. Mit Description Schemes kann man sehr komplexe Beschreibungen erzeugen, die sich neben Deskriptoren auch aus mehreren Description Schemes zusammensetzen lassen.

Abbildung 2 ist ein Beispiel für ein Description Scheme. DS „2“ könnte z.B. eine einfache Beschreibungsstruktur für ein Bewerbungsvideo darstellen, die Deskriptoren „A“ und „B“ könnten Ort und Zeit beschreiben, und das Description Scheme „1“ könnte die bewerbende Person mit Name und Alter als Deskriptoren „C“ und „D“ darstellen.

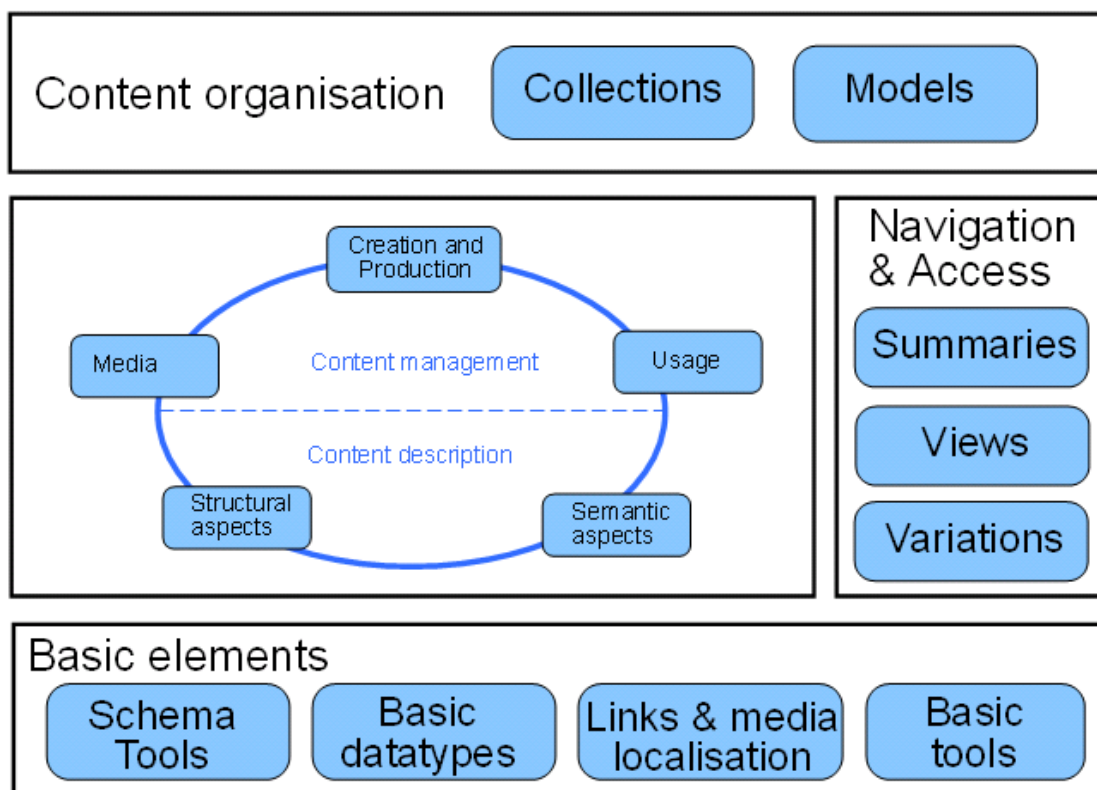


Abbildung 3: Media Description Schemes [Martinez, 2004]

2.4. Media Description Schemes

Die Media Description Schemes sind eine Sammlung von Deskriptoren und Description Schemes. Sie bilden eine umfangreiche Bibliothek, in der alle grundlegenden und nötigen Bausteine enthalten sind, um Audio- und Videoeigenschaften umfassend zu beschreiben. Wie man in Abbildung 3 sehen kann, gliedert sich die Bibliothek in fünf Teile.

Die *Basic Elements* sind eine Reihe von grundlegende Basisbausteinen, die erweiterte Datentypen und mathematische Strukturen wie z.B. Vektoren und Matrizen enthalten. Desweiteren enthalten die Basiselemente Konstrukte, um Verknüpfungen zwischen Description Schemes zu realisieren und um Teilinhalte zu lokalisieren.

Bei *Content Management* stellt MPEG-7 Description Schemes für die Verwaltung von Multimediadokumenten zur Verfügung. Von Informationen zur Erzeugung und Produktion, zum Quellmedium bis zur späteren Verwendung und zu den Verwertungsrechten stehen hier die notwendigen Description Schemes zur Verfügung. Semantische Aspekte wie z.B. die Fragen nach dem „Wer?“, „Wie?“, „Warum?“ und und strukturelle Aspekte wie z.B. Farbe und Textur decken die Description Schemes des Bereichs Content Description ab.

Die *Navigation and Access* Description Schemes dienen zum Beschreiben von Zusammenfassungen, verschiedenen Partitionen und Sichten wie z.B. Kapitel oder mehre Audiospuren und Kameraeinstellungen in einem Video.

Der letzte Teil der Media Description Schemes ist die *Content Organisation*. Hier werden Description Schemes bereitgestellt, mit denen man Audio- und Videoinhalte, Segmente, Ereignisse und Objekte mit gemeinsamen Eigenschaften beschreiben kann. Desweiteren kann man mit verschiedenen Modellen und Statistiken diese Kollektionen noch genauer beschreiben, um die gleichen Eigenschaften der dazugehörigen Elemente und Daten zu charakterisieren.

3. Ausgewählte Deskriptoren

Im Weiteren werden nun exemplarisch vier ausgewählte visuelle Deskriptoren näher beschrieben.

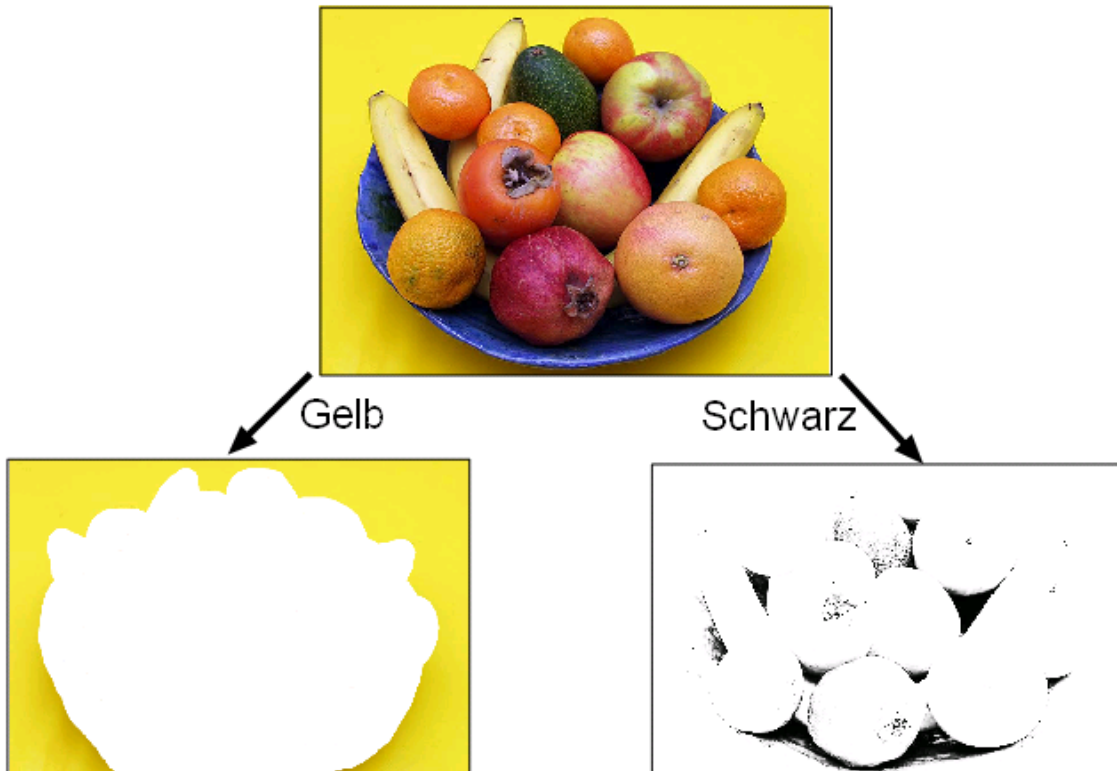


Abbildung 4: Color Structure Descriptor

3.1. Color Structure Descriptor

Der Color Structure Descriptor beschreibt die Kompaktheit einer einzelnen Farbe. Wenn man, wie in Abbildung 4 zu sehen, die Farben Gelb und Schwarz aus dem ursprünglichen Bild filtert, kann man sofort erkennen, dass Gelb relativ kompakt ist und eine zusammenhängende Fläche bildet, wobei Schwarz sehr im Bild stark verstreut ist. Anhand dieser Eigenschaft kann man sehr schnell zwei Bilder vergleichen und große Unterschiede sofort erkennen, ideal für eine grobe Vorauswahl bei einer Bildsuche.

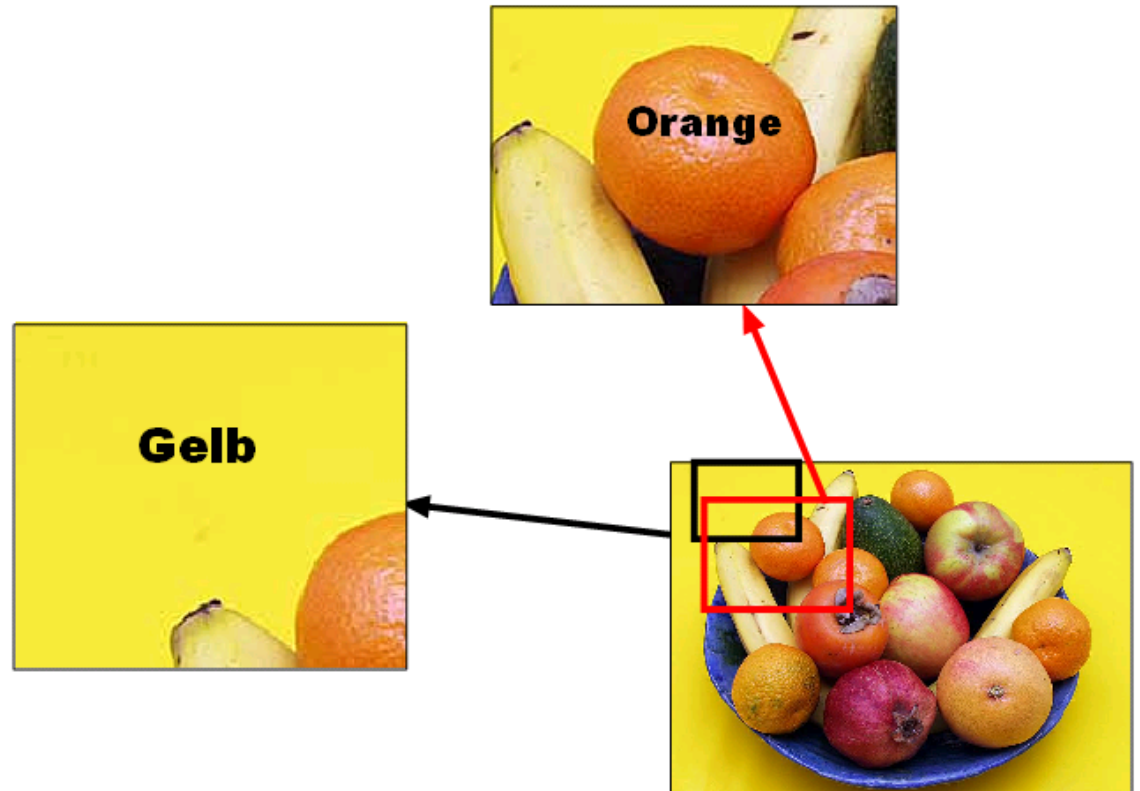


Abbildung 5: Dominant Color Structure

3.2. Dominant Color Descriptor

Eine andere Möglichkeit, um ein Bild oder einen Bildausschnitt zu charakterisieren, ist die Suche nach der am stärksten vertretenen Farbe. Diese Eigenschaft spiegelt der Dominant Color Descriptor wieder. In Abbildung 5 lässt sich auf einen Blick erkennen, dass die beiden Bildausschnitte nicht sehr ähnlich sind.

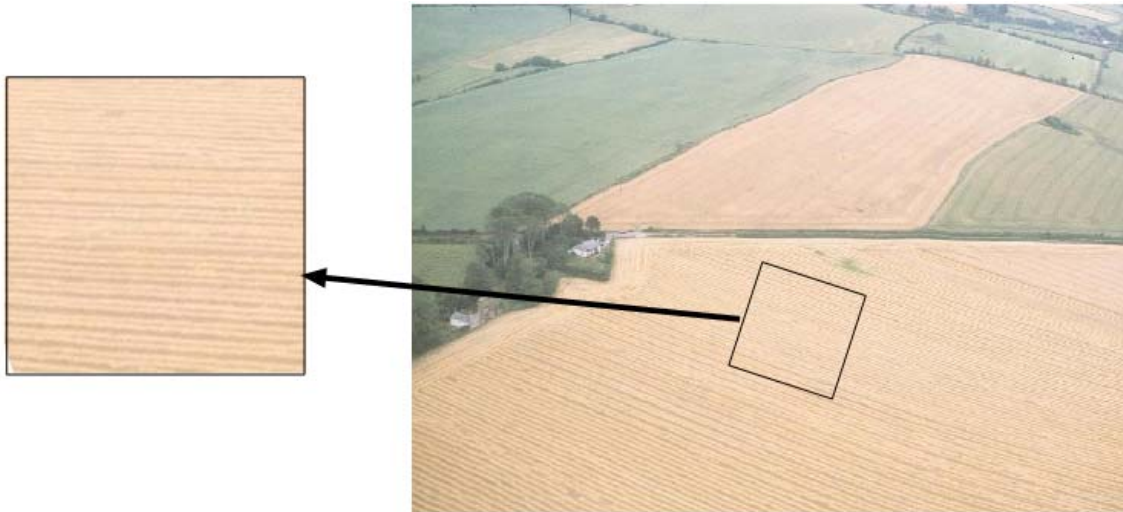


Abbildung 6: Homogeneous Texture Descriptor

3.3. Homogeneous Texture Descriptor

Wenn man anhand eines Musters z.B. eine Luftaufnahme wie in Abbildung 6 auf Ähnlichkeit hin analysieren möchte, kann man dazu den Homogeneous Texture Descriptor verwenden, der eine Textur beschreibt. Durch Filtern, Drehen und Skalieren können Entfernung und Perspektive einer solchen Luftaufnahme an die Textur angeglichen werden und ein Vergleich vorgenommen werden, da man damit eine gemeinsame Vergleichsbasis erhält.

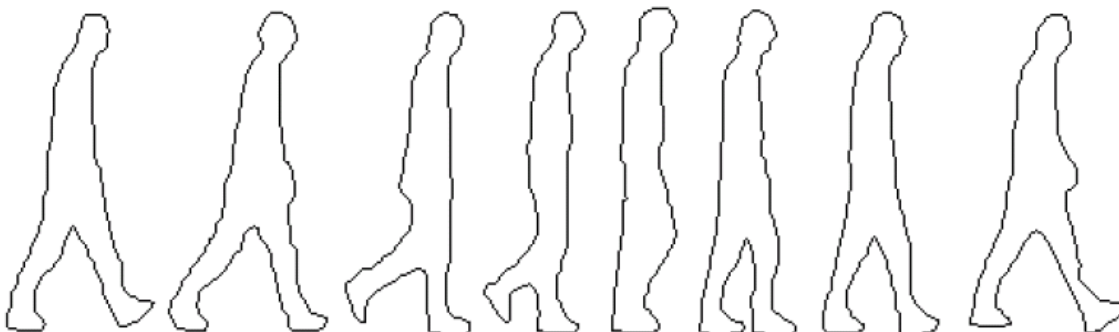


Abbildung 7: Kontur eines gehenden Menschen [Kopf]

3.4. Contour Shape Descriptor

In der automatischen Videoanalyse ist die Objekterkennung ein wichtiger Bestandteil. Dadurch ist es möglich, für eine umfangreiche Multimediabibliothek einen Index zu erstellen und eine semantische Suche darin zu ermöglichen. In der Wahrnehmungspsychologie fand man heraus, dass das menschliche Gehirn dreidimensionale Objekte anhand ihrer zweidimensionalen Kontur (Abbildung 7) erkennen kann. Mithilfe einer automatischen Klassifikation wird es ermöglicht, die Objektklasse zu bestimmen und so z.B. Videomaterial aus unterschiedlichen Genres zu analysieren [Kopf].

MPEG-7 bietet hierfür den Contour Shape Deskriptor, für den die äußere Kontur eines Objektes durch eine vorher festgelegte Anzahl von Wertepaaren in einem zweidimensionalen Raster beschrieben wird.

Um zwei Konturen auf Ähnlichkeit hin zu untersuchen wendet man das Verfahren des Krümmungsbasierter Skalenraums an. Die Idee ist, die markantesten Krümmungen einer Kontur als Vergleichspunkte zu wählen.

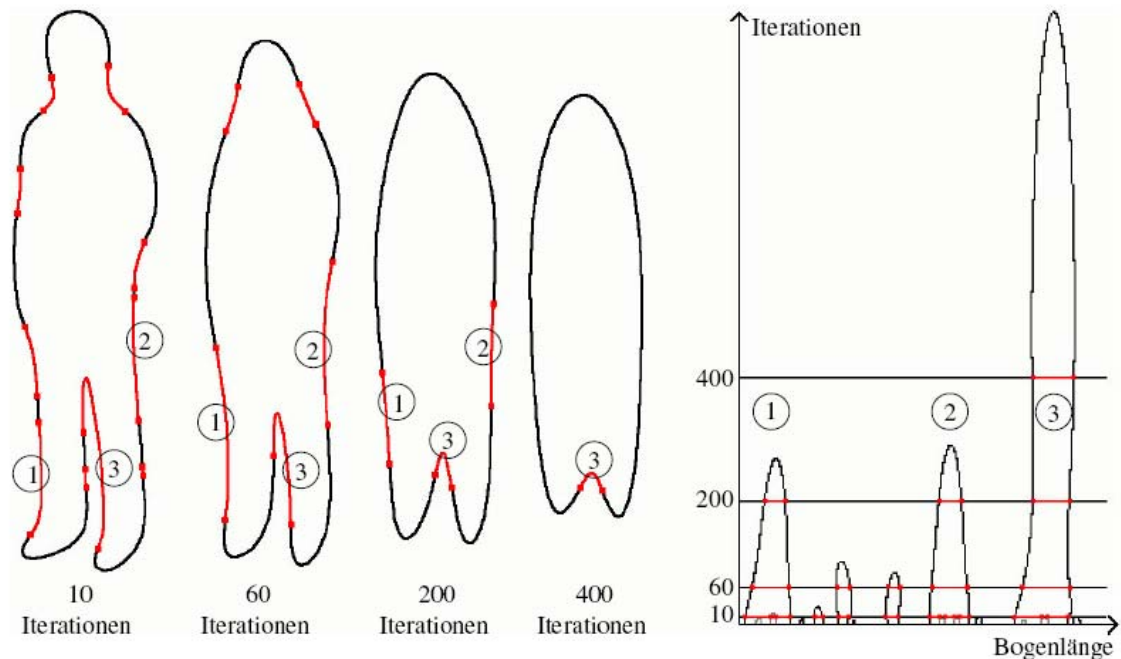


Abbildung 8: Krümmungsbasierter Skalenraum [Kopf]

In Abbildung 8 kann man auf der linken Seite die Kontur eines Menschen sehen, bei der nach wenigen Iterationen von Glättungen mit einem Gaussfilter die kleinen Bögen

verschwunden sind und nur noch drei Markante die Glättungen überstehen. Die Grafik rechts stellt genau diese Informationen in einem Koordinatensystem dar. Nach rechts trägt man die Position der Wendepunkte der Bögen auf der Kontur ab, wobei der Startpunkt dafür willkürlich gewählt wird. Die Höhe der dargestellten Bögen zeigt die Änderung der Wendepunkte nach den jeweiligen Glättungsiterationen. Bei Bogen Nr. 3 kann man sehr gut erkennen, dass er der markanteste und ausgeprägteste der Kontur ist. Seine Wendepunkte liegen im Vergleich zu den beiden anderen Bögen noch nicht aufeinander (d.h. der Bogen ist nicht mehr existent), sondern der Abstand ihrer Position auf der Kontur hat sich nur ein wenig verringert. Beim Vergleichsvorgang wählt man nun die beiden höchsten Bögen der Konturen aus und berechnet die Differenz ihrer Höhen und Positionen, die einen festgelegten Grenzwert nicht überschreiten dürfen, da sonst der Vergleich aufgrund mangelnder Ähnlichkeit abgebrochen wird. Aus der Summe der Differenzen aller Bögen wird dann ein Wert berechnet, der die Unähnlichkeit der beiden Konturen beschreibt und somit als Distanzmaß dient zum Vergleich dient.

3.5. Distanzmaß

Um Multimediadaten vergleichen und eine Suche anhand von MPEG-7 Deskriptoren ermöglichen zu können, benötigt man ein Distanzmaß, anhand dessen man eine Ähnlichkeit festzutellen und bewerten kann. Da sehr viele der Deskriptoren als Vektor dargestellt werden können ist die euklidische Distanz ein geeignetes Maß. Sie wird im Standard empfohlen, da sie schnell und einfach berechnet werden kann.

$$d(x, y) = \sum_{i=1}^n (x_i - y_i)^2$$

4. Fazit

Durch ein einheitliches Format für Metadaten ist der Grundstein für eine breite Palette von Anwendungen gelegt, die diese auslesen und weiterverarbeiten können. Sind die entsprechende Deskriptoren vorhanden, kann man z.B. bei der multimodalen Suche nicht nur nach einem Stichwort sondern auch anhand einer handgemalten Skizze oder einer Stimmenaufnahme suchen. Bei der automatischen Videoüberwachung hätte man die Möglichkeit, gefährliche Objekte wie z.B. ein Messer oder andere Waffen zu erkennen, und könnte einen Alarm auslösen. Für blinde Menschen könnte z.B. eine computergenerierte Stimme eine wortlose Szene in einem Videofilm anhand der MPEG-7 Beschreibungen der dargestellten Objekte zusammenfassen.

MPEG-7 bietet sehr umfangreiche Möglichkeiten bei der Beschreibung von Multimediadaten und ist auch für zukünftige Anwendungen offen, da der Standard nur die Form der Beschreibungsstruktur festlegt und erweiterbar ist. Aber gerade der große Umfang und die Komplexität von MPEG-7 ist wohl mit ein Grund dafür, dass der Standard noch keine weitläufige Anwendung gefunden hat.

5. Literaturverzeichnis

Marcel Richter: „MPEG-7: Ein Standard zur Beschreibung von Multimedia-Inhalten“, Universität Bielefeld, 2003

Michael Hönyck: „Multimediabeschreibung mittels Beschreibungssprache: MPEG-7“, Institut für Nachrichtentechnik Aachen, 2004

Jose M. Martinez: „MPEG-7 Overview“, International Organisation for Standardisation, 2004

H. Kosch, Informatiklexikon: „MPEG-7“, Informatiklexikon, Gesellschaft für Informatik E.V., 2003

Stephan Kopf: „Computergestützte Inhaltsanalyse von digitalen Videoarchiven, Kapitel 5: Klassifikation von Objekten“, Universität Mannheim (noch nicht veröffentlicht).