

Annotationssysteme für Metadaten

Seminararbeit

vorgelegt am
Lehrstuhl für Praktische Informatik IV
Prof. Dr.-Ing. W. Effelsberg
Universität Mannheim
im
Juni 2006

von
Aleksandra Dimitrijević

im Rahmen des
Seminars „Bridging the Semantic Gap“
unter der Leitung von
Prof. Dr. W. Effelsberg
Betreuer
Dipl.-Wirtsch.-Inf. Stephan Kopf

Inhalt

Inhalt	3
Abbildungsverzeichnis	4
Abkürzungsverzeichnis	5
1. Einleitung	6
2. Begriff der Annotation	7
3. Annotation von Multimediadaten	9
3.1. Besonderheiten von Annotationssystemen	9
3.2. Herausforderungen	10
3.3. Systeme in Überblick	11
3.3.1. Architektur	11
3.3.2. Beispiele	13
4. Schlusswort	19
Literaturverzeichnis	20

Abbildungsverzeichnis

Abb. 1. Grafische Oberfläche des VideoAnnEx Tools	12
Abb. 2. Modell-Training-Prozess bei dem Video Automatic Labeling System	13
Abb. 3. Labeling-Prozess bei MPEG-7 Video Automatic Labeling System	13
Abb. 4. Grafische Oberfläche des MultiMedia MultiModal Annotation Tools	14
Abb. 5. Grafische Oberfläche des iVAS Tools	16
Abb. 6. Ergebnisse von Soccer highlight automatic video annotation	17
Abb. 7. Durch Bilder erweiterte Ontologien	17

Abkürzungsverzeichnis

DC	<i>Dublin Core</i>
MARC	<i>MAchine Readable Catalogue</i>
MPEG	<i>Moving Picture Experts Group</i>
ISO	<i>International Standard Organization</i>
XML	<i>Extensible Markup Language</i>
W3C	<i>World Wide Web Consortium</i>
EXIF	<i>Exchangeable Image File Format</i>
JPEG	<i>Joint Photographic Experts Group</i>
JEIDA	<i>Japan Electronic Industry Development Association</i>
IBM	<i>International Business Machines</i>
iVAS	<i>intelligent Video Annotation Server</i>

1. Einleitung

Computer werden eingesetzt, um Informationen zu speichern, zu bearbeiten, zu kombinieren und zu produzieren, wodurch eine enorme Menge von Daten entsteht. Das Ziel ist es, die Arbeit des Computers so effektiv wie möglich, in Bezug auf Sortieren und Organisieren dieser großen Datenmenge, zu machen. Heutzutage ist das Internet ein sehr wichtiges Instrument in unserer Wissensgesellschaft. Es repräsentiert eine große Quelle von Informationen, Daten und Diensten, die ständig wächst. Das Problem besteht darin, in dieser enormen Menge von Dateien, die richtigen Informationen zu finden. Die Lösung des Problems sieht so aus, dass der Computer mit in diese Suche beteiligt wird. Bislang basiert die Suche nach Informationen auf dem Inhalt der Dateien und nicht auf der eigentlichen Bedeutung von Gleichen. Momentan sind Inhalte im Internet nur maschinenlesbar, jedoch gleichzeitig von Computer nicht verstehbar. Die Metadaten werden eingesetzt, um dieses Problem zu lösen. *Metadaten* dienen zur Identifikation, Beschreibung und Verwaltung von Informationsobjekten. Es stellt sich die Frage, wie der Annotierungsprozess durchgeführt werden soll.

Die semantische Annotation von Dateien ist ein sehr wichtiger Prozess, bei dessen Durchführung gleichzeitig viele Probleme entstehen, besonders wenn es um Multimedia-dateien geht.

Diese Arbeit beschäftigt sich mit Annotationsystemen für Multimediadateien. Es werden die Besonderheiten und Probleme bei der Annotation erläutert. Des Weiteren erfolgt eine ausführliche Betrachtung unterschiedlicher Annotationssysteme. Zum einen wird die Architektur der Systeme beschrieben, zum anderen werden ausgewählte Annotationssysteme vorgestellt.

2. Begriff der Annotation

Das Ziel des Semantischen Webs ist es, Daten so umzuwandeln, dass der Computer den Kontext, das Format und die Bedeutung des Inhaltes versteht. Die Grundidee des Semantischen Webs basiert auf einer Verwendung von *Metadaten*, die die Bedeutung (Semantik) von Informationen für den Computer bereitstellt. Mit Hilfe von Metadaten werden die Programme auf der Basis von Bedeutung operieren können. „Metadaten sind strukturierte Informationen über Informationsobjekte. Sie dienen zur Identifikation, Beschreibung und Verwaltung von Informationsobjekten. Sie dienen als Ordnungssysteme, die Informationsobjekte unter bestimmten Kriterien auffindbar machen - in diesem Sinne dienen sie der Verbesserung des Information Retrieval.“ [1]

Unter *Semantischer Annotation* (engl. semantic annotation) versteht man das Hinzufügen von semantischen Metadaten in Dokumente. Diese beschreiben den Inhalt eines Dokuments in maschinen-verarbeitbarer Form. Bei den Dokumenten kann es sich um Informationsquellen wie z.B. Web-Seiten, Bilder, Audio, Video Dateien handeln.

Das Ziel der Forschung und Entwicklung im Bereich der Metadaten ist die Interoperabilität verschiedener Metadatenformate und verschiedener technischer Systeme. „Als Interoperabilität bezeichnet man die Fähigkeit zur Zusammenarbeit von verschiedenen Systemen, Techniken oder Organisationen. Dazu ist in der Regel die Einhaltung gemeinsamer Standards notwendig“ [2]. Deswegen sind Standards für Syntax und Semantik nötig, die für entsprechende Dateiformate entwickelt wurden.

Zu den bekannten Standards zählen: Dublin Core (DC), Text Encoding Initiative, Encoding Archive Description, Government Locator Service und MARC (MACHINE READABLE Catalogue).

MPEG-7 ist ein Metadaten-Standard für Multimediadaten aller Art. Es wurde von der Gruppe *Moving Picture Experts Group* der ISO entwickelt. Mit MPEG-7 wurde der Versuch gemacht, einen Standard zu entwerfen, der anwendungsunabhängig ist und das gesamte Spektrum des Multimedia-Bereichs abdeckt. Zudem bietet MPEG-7 flexible Strukturierungsmöglichkeiten und baut auf dem allgemein anerkannten XML-Schema auf [6].

Metadaten kann man manuell oder durch einen automatischen Prozess erstellen. Da die manuelle Annotation viele Nachteile hat, werden Systeme entwickelt, die diesen Prozess automatisieren und erleichtern.

Nachteile von manueller Annotation sind:

- Durch die Bereitstellung der Personalkosten ist die manuelle Annotation teuer.

- Manuelle Annotation ist sehr zeitaufwändig, da, bevor man eine Datei beschreiben kann, sie zuerst analysiert werden muss, d.h. Video Dateien müssen angesehen werden, Audio Dateien angehört werden. Danach kommt die eigentliche Annotation.
- Manuelle Annotation kann sehr mühsam sein, d.h. die Person die diesen Prozess durchführt muss stets konzentriert und motiviert sein, des weiteren sollte sie über genug Wissen zum Thema haben, damit der Prozess richtig durchgeführt werden kann. Alles andere wirkt sich negativ auf den Annotierungsprozess aus.
- Manuelle Annotation sollte nicht subjektiv sein, d.h. die Person sollte eine objektive Beschreibung zum Thema geben. Die Beschreibungen können von psychischen, intellektuellen, sozialen Faktoren abhängig sein.

Vorteil der manuellen Annotation ist:

- Beschreibungen die manuell erstellt werden, können ein richtiges Abstraktionslevel haben. Die Person, die die Aufgabe hat, die Metadaten zu erstellen, kann sehr gut die Bedeutung von Dateien verstehen. So kann Semantik korrekt beschrieben werden.

Im Annotierungsprozess sollten alle relevante Informationen über bestimmte Ressourcen eingegeben werden. Folgende Typen von Metadaten können unterschieden werden:

1. Metadaten zur Identifikation und zu Nachweiszwecken ("resource discovery"), auch administrative Daten.
2. Metadaten für Zugangsbedingungen sowie Nutzungs- und Beschaffungskonditionen („terms and conditions“).
3. Metadaten zu strukturellen Aspekten („structure“).
4. Metadaten zum Kontext („context“).
5. Metadaten zum Inhalt („content“).
6. Metadaten zur Nutzungs- und Verwendungsgeschichte („use history“). [3]

Man sieht gleich, dass die Erstellung von Metadaten zum Kontext oder Inhalt, am kompliziertesten ist, sowohl bei manuell Annotation, als auch bei Annotationssysteme.

Im Verlauf dieser Arbeit wird ausschließlich die automatische Annotation behandelt. Da dieser Prozess ein zeitaufwändiger Prozess ist, wurden verschiedene Tools und Systeme entwickelt, die diesen erleichtern. Großer Aufwand entsteht besonders beim Annotierungsprozess von Multimedia Dateien.

3. Annotation von Multimediadaten

Bevor man sich mit Annotationssystemen für Multimedia Dateien beschäftigt, sollte man zuerst über die Eigenschaften von Multimedia Dateien bescheid wissen. Dazu werden zunächst die Besonderheiten von Annotierungsprozessen von Multimediadateien erläutern, sowie die Probleme die dabei entstehen.

3.1. Besonderheiten von Annotationssystemen

Die Annotierung von Webseiten und Texten kann auch wie die Suche im Web, textbasiert sein. Dies bedeutet, dass die Erstellung von Metadaten für Webseiten, durch eine Textanalyse, d.h. eine Analyse des Inhaltes der Seite, unterstützt wird. So können verschiedene Werkzeuge entwickelt werden, die z.B. Vorschläge zu ausgewählten Deskriptoren machen. Bei automatischer Indexierung kann durch die Deskriptoren der semantische Inhalt festgelegt werden. Auch für dynamisch generierte Seiten ist es möglich ein Annotierungsmuster festzulegen.

Ähnlich wie bei der Indizierung von Textdaten soll das in verschiedenen Multimediatypen wie Video, Audio, Bilder und Animationsdateien enthaltene Wissen erschlossen und gespeichert werden.

Multimediadateien sind im Vergleich zu reinen Textdateien sehr komplex, da die Inhalte in Multimediadateien örtliche und zeitliche Dimension haben. Auch haben Forschungsergebnisse gezeigt, dass die Wahrnehmung von Musik durch viele Kriterien beeinflusst wird, die nicht bewusst hörbar sind [4]. Das gilt für alle Multimediadaten, besonders für Video Dateien.

Für Annotation bedeutet dies, dass die Metadaten nicht so leicht erfasst werden können. Wenn es um manuelle Annotation geht, muss zuerst eine visuelle Analyse von Bildern, Videodateien oder eine Audioanalyse (anhören) von Audiodateien durchgeführt werden, bevor eine Beschreibung für entsprechendes Material erstellt werden kann. Dieser Prozess kann sehr zeitaufwändig und mühsam sein. Es werden Tools benötigen, die diese Prozesse erleichtern, automatisieren und vereinfachen. Schnell wird klar, dass diese Aufgabe (eine Datei zu analysieren und den Inhalt, die Attribute, die Bedeutung und Beziehungen zu verstehen) eine sehr große Herausforderung für ein Annotationsystem ist.

3.2. Herausforderungen

Da es im Internet schon eine große Menge von Dateien gibt, die nicht annotiert sind, stellt sich die Frage, wie bei diesen Dateien die Annotierung durchgeführt werden soll und des weiteren, wie man die Leute dazu motivieren kann die Annotation zu machen und das Annotationstools zu benutzen. W3 Konsortium berücksichtigt folgende fünf Punkte, die Bereitschaft beeinflussen, Dateien zu annotieren und zugleich die Qualität der Annotation zu gewährleisten.

1. Annotation während der Produktionsphase versus Annotation nach der Produktionsphase

Normalerweise ist es immer besser den Annotierungsprozess möglichst früh durchzuführen. In Prinzip sind viele wichtige Informationen während der Produktion von Dateien vorhanden. z.B. Uhrzeit und Datum, Blendeneinstellung (lens settings), und die weitere EXIF Metadaten, die zu den JPEG Bildern automatisch durch manche digitale Kameras hinzugefügt werden. Das **Exchangeable Image File Format (EXIF)** ist ein Standard der Japan Electronic Industry Development Association (JEIDA) für das Dateiformat, in dem moderne Digitalkameras ihre Daten speichern [5]. Das Annotieren während der Produktionsphase ist viel günstiger und erbringt bessere Ergebnisse als das Hinzufügen der Metadaten in einem späteren Prozess.

2. Allgemeine versus spezifische Annotation

Annotation von Bildern und Videos ohne ein bestimmtes Ziel, einen Kontext oder eine konkrete Aufgabe, führt häufig zu schlechten Ergebnissen und hohen Kosten. So können die Bilder z.B. bei der Ausführung eines Annotationsprozesses, mit einer ungenügenden Information für den Annotationsprozess annotiert werden. Die Neubearbeitung des Annotationsprozesses ist dann eine unvermeidliche und teure Lösung. Auf der anderen Seite, können Beschreibungen, die zu spezifisch sind, nicht wieder verwendet werden.

3. Manuelle versus automatische Annotation

Die manuelle Annotation hat den Vorteil, dass die Semantik der Daten korrekt beschrieben wird (das Abstraktionslevel ist passend). Andererseits ist dieses Verfahren sehr zeitaufwändig und teuer und kann oft subjektiv sein. Automatische Annotierung ist dagegen nicht teuer, ist schnell und systematisch, liefert im Allgemeinen jedoch nur die Low-Level Beschreibung. Diese Unterschied zwischen Low-Level Descriptoren und High-Level Beschreibungen nennt man Semantic Gap.

4. Verschiedene Typen von Metadaten

Zum einen unterscheidet man zwischen Metadaten die die Eigenschaften des Bildes beschreiben und zum anderen Metadaten die die Inhalte beschreiben (z.B. die Eigenschaften des Objekts, der Person oder des Geschehens, die das Bild darstellt). Zur ersten Kategorie gehören die Angaben des Titels, Autors, die Auflösung des Bildes, Ausgabezeit, etc. Mit der zweiten Kategorie wird beschrieben, was auf dem Bild dargestellt ist. Bezüglich der zweiten Kategorie muss man zwischen domänenspezifischen Vokabular und domänenunabhängigen Vokabular unterscheiden (objektive Beobachtungen und subjektive Interpretationen).

5. Mangel von syntaktischer und semantischer Interoperabilität

Bei dem Annotationsprozess von Bildern können viele unterschiedliche Dateiformate und Anwendungen genutzt werden. Die Bearbeitung der Dateien, die mit Hilfe eines anderen Tools erstellt worden sind, wird oft durch den Mangel der Interoperabilität eingeschränkt. Folgendes Beispiel könnte auftreten:

- Bestimmte Tools nutzen eine andere Syntax für ihre Dateiformate. Als Folge können die durch diese Systeme geschaffene Annotation nicht immer gelesen werden.
- Bestimmte Tools liefern eine andere Bedeutung für das gleiche Objekte.[6]

3.3. Systeme in Überblick

3.3.1. Architektur

Das Problem des Unterschieds zwischen dem Inhalt der Multimediadateien und dem „Wissen“ des Computers wird mit Hilfe der Annotationssysteme verringert. Auf Grund der unterschiedlichen Kriterien kann man die Annotationssysteme in verschiedene Gruppen teilen. Folgende Kriterien sind bei der Architektur von Annotationssystemen wichtig:

- **Kollaborative oder Individuelle Systeme:** Damit ist gemeint, ob ein System für einen einzelnen Benutzer oder für die ganze Gruppe gedacht ist.
- **Annotation level:** Es ist einmal von der Implementierung der Lexikas abhängig und andererseits vom Schreiben freier Texte. Wenn man die Ontologien benutzt, spricht man von einem höheren Annotations Niveau, weil dann die Semantik auf einem formalen Weg beschrieben wird.

- **Niveau von Automatisierung:** Hierbei wird festgelegt, wie hoch der Automatisierungsgrad des lernenden Systems ist.

Die Architektur der Annotationssysteme hängt zum einen davon ab, welcher Gruppe dieses System angehört und zum anderen vom erreichten Niveau der verschiedenen Kriterien. Die Gemeinsamkeit der Systeme, besteht darin, dass jedes dieser Systeme aus zwei Modulen besteht, die die Ressource analysieren und den Annotierungsprozess durchführen.

Das erste Modul soll die Annotierungsprozesse erleichtern. Es implementiert bei der Video Analyse folgende Algorithmen: Szenenerkennung, Erkennung von Kamerabewegungen, Farbhistogramme, Texterkennung, „Action“-Intensität, Gesichtserkennung, Objekterkennung, etc. Bei der Audio Analyse benutzt man folgende Algorithmen: Segmentierung, Sprecher-Erkennung, Hauptmelodie-Erkennung, Erkennung von Schlagzeug-Schlägen...

Ziel ist es, die Datei zu analysieren und den eigentlichen Annotierungsprozess vorzubereiten, indem man z.B. bei den Videos die Schnitte erkennt. Dieses Modul liefert auch Low-Level Beschreibungen, aus denen XML-Dateien generiert werden können. Mit Low-Level ist eine niedrigere Abstraktionsebene gemeint. Auf dieser Ebene werden Objekte und Segmente, sowie deren grundlegende Eigenschaften zugeordnet. Diese Eigenschaften können z.B. Farben, Umrisse, Töne und Bewegungen sein. Es sind Eigenschaften, die relativ einfach aus den Multimedia-Daten errechnet werden können [7]. Nach der Analyse von Dateien werden die Informationen ermittelt, die eine Basis für semantische Aussagen über den Inhalt präsentieren.

Im zweiten Modul werden die Annotationprozesse durchgeführt, wobei es sich hierbei meistens um eine manuelle Annotation handelt.

3.3.2. Beispiele

1. IBM MPEG-7 Annotation Tool

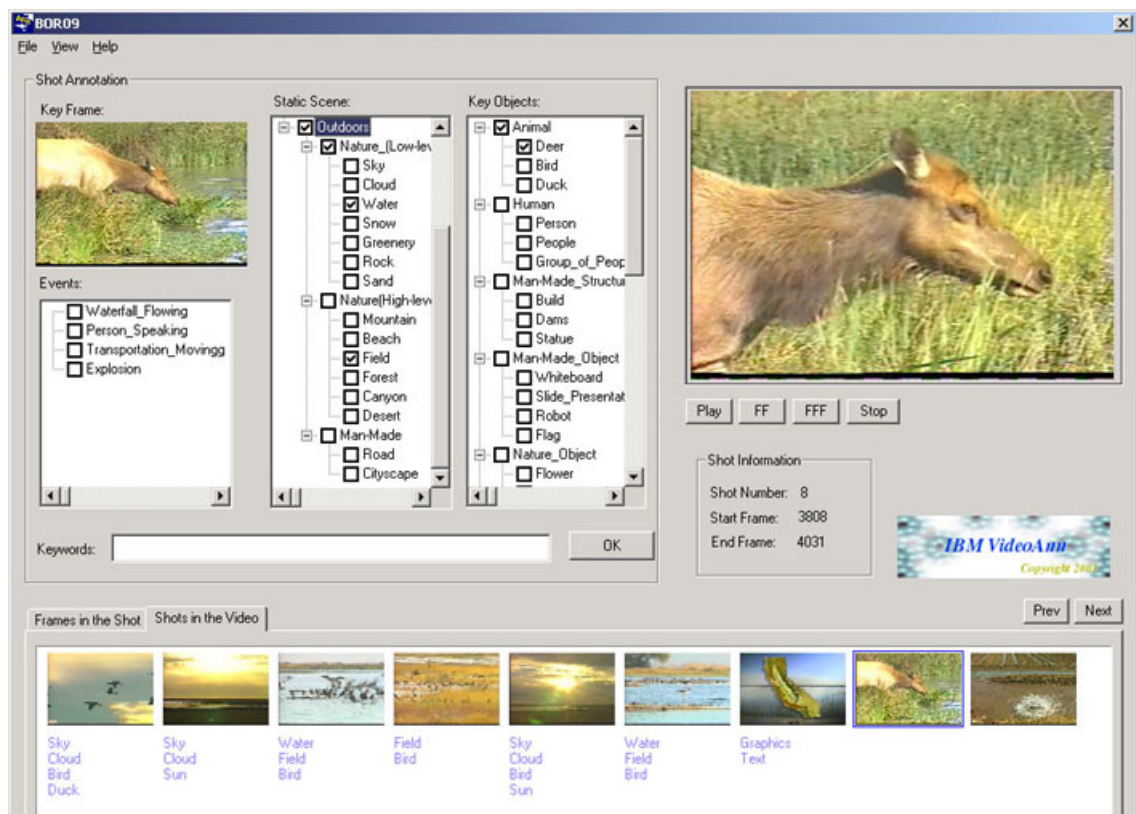
IBM hat ein MPEG-7 Annotationssystem VideoAnnEx entwickelt. Das Annotationssystem „VideoAnnEx“ unterstützt die Annotation von Videos mit MPEG-7 Metadaten. Mit diesem Tool kann man Annotation Lexikas bilden, speichern, downloaden und updaten. Die Videos werden in Schnitte unterteilt. Für jeden Schnitt kann man statische Szenen (static scene), Schlüsselobjekte (key objects) und Ereignisse (events), beschreiben. Die Beschreibungen werden als MPEG-7 Beschreibungen in einer XML-Datei gespeichert.

Wenn ein Video geöffnet ist, wird es durch Szenenerkennung in Schnitte unterteilt. Anschließend, können für diese Schnitte Informationen in ein MPEG-7 Schematas gespeichert werden. Wenn ein Annotationslexikon existiert, wird es benutzt, ansonsten kommt das vorgegebene Lexikon zur Anwendung.

VideoAnnEx ist in vier graphische Teile unterteilt:

1. Video Playback mit Schnitt Informationen.
2. Schnitt Annotation mit Annotationslexikon.
3. Sichtleiste mit allen Videoschnitten.
4. Annotation von Regionen in einem bestimmten Schnitt.

Abb. 1. Graphische Oberfläche des VideoAnnEx Tools

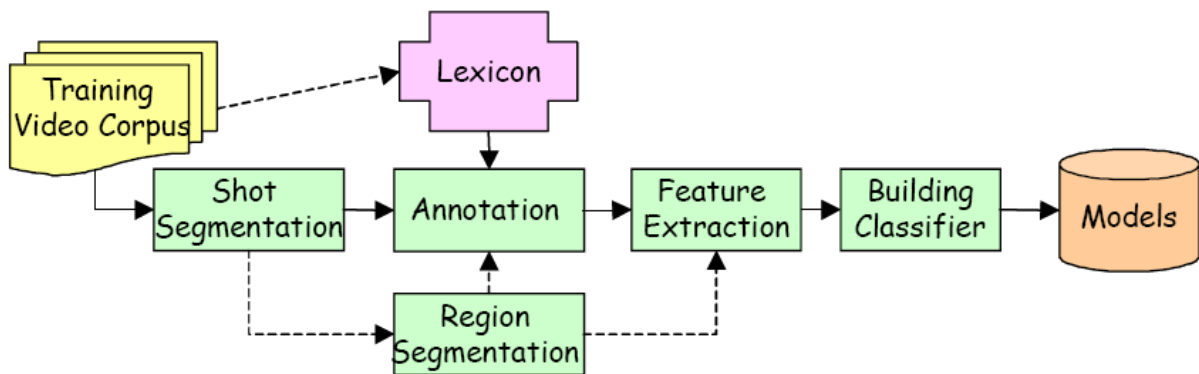


2. MPEG-7 Video Automatic Labeling System [9]

Dieses System ist ein lernendes System zum automatischen Annotieren von MPEG-1 Videos, wobei die Ausgang eine MPEG-7 XML Datei ist. Das System besteht aus zwei Teilen:

- Modell – Trainings - Prozess
- Labeling- Prozess

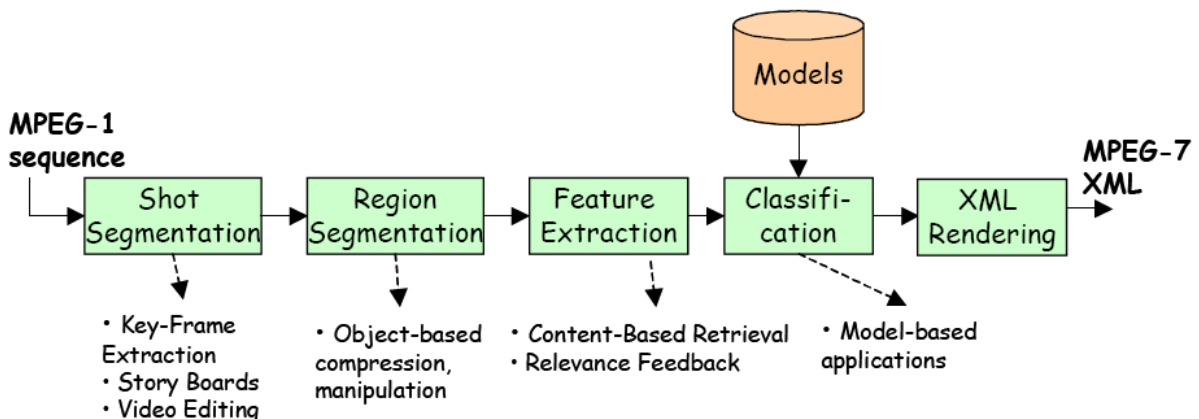
Abb. 2. Modell-Training-Prozess bei MPEG-7 Video Automatic Labeling System



Quelle: [9]

Im ersten Schritt des („model training process“), wird beim Video eine Schnitterkennung durchgeführt. Danach werden jedem Schnitt oder jeder Region Beschreibungen zugeteilt (hier wird das Video Ann Ex Tool benutzt). Das „Feature Extraction Modul“ extrahiert visuelle Merkmale in zeitliche und örtliche Dimension. Am Ende des lernenden Moduls werden Modelle (Grundkonzepte) gebildet, die verschiedenen Erscheinungen entsprechen, z.B., im Freien, Innen, Gesicht, Person, Landschaft, Text, Rede, Instrumental, Ton, Monolog.

Abb. 3. Labeling-Prozess bei MPEG-7 Video Automatic Labeling System



Quelle: [9]

In zweiten Schritt dem „labelling process“, sind die ersten drei Module mit dem „model training process“ identisch. Wenn die Features extrahiert sind, testet das Klassifikationsmodul die Wichtigkeit der Schnitte, indem es sie mit den Modellen (die in der ersten Phase gebildet wurden) vergleicht. Hierzu werden die „Support Vector Machines“ Klasifikation und „Fusions-Methoden“ verwendet. Das Ergebnis ist der Vertrauens Wert jedes Konzeptes.

Am Ende werden die MPEG-7 XML Dateien generiert. Das gilt jedoch nur für Konzepte mit einem großen Vertrauenswert, d.h. nur für Videoschnitte (oder Regionen), bei denen die Konzepte erkannt werden.

3. M4Note, MultiMedia MultiModal Annotation Tool [10]

Das wichtigste Merkmal dieses Systems ist die Benutzerfreundlichkeit. Bei der Implementierung war vor allem wichtig, die Benutzerfreundlichkeit zu erhöhen. Bei der Annotierung wird „pen-based electronic ink“ und „voice recognition“ verwendet. Mit der Benutzung von „handwriting“ kann man auch bestimmte Symbole definieren, die dem entsprechenden Benutzer zugeteilt werden.

Abb. 4. Graphische Oberfläsche des MultiMedia MultiModal Annotation Tool



Quelle: [10]

4. iVAS: Web-based Video Annotation System [11]

iVAS (intelligent Video Annotation Server) ist ein Web basierendes Annotations-system. Die Idee besteht darin, dass den Annotierungsprozess auch normale Internet-Nutzer durchführen können. Somit basiert das System auf Web-Browser, d.h. der Browser bietet als Erweiterung die Annotationsfunktion. Zuerst wird das Video analysiert und danach in Schnitte unterteilt. Hier werden auch Informationen zu Farbhistogrammen zusammengestellt. Die erstellten Annotationen werden im Internet in Annotation-XML-Datenbanken gespeichert.

Dieses System unterstützt 3 Arten von Annotationen:

1. Text-Annotation:

Bei der Text Annotation werden für jeden Schnitt Kommentare und Beschreibungen erstellt. Man kann zwischen verschiedenen Typen von Kommentaren wählen: Name, Situation, Beschreibung, Kommentar, etc.

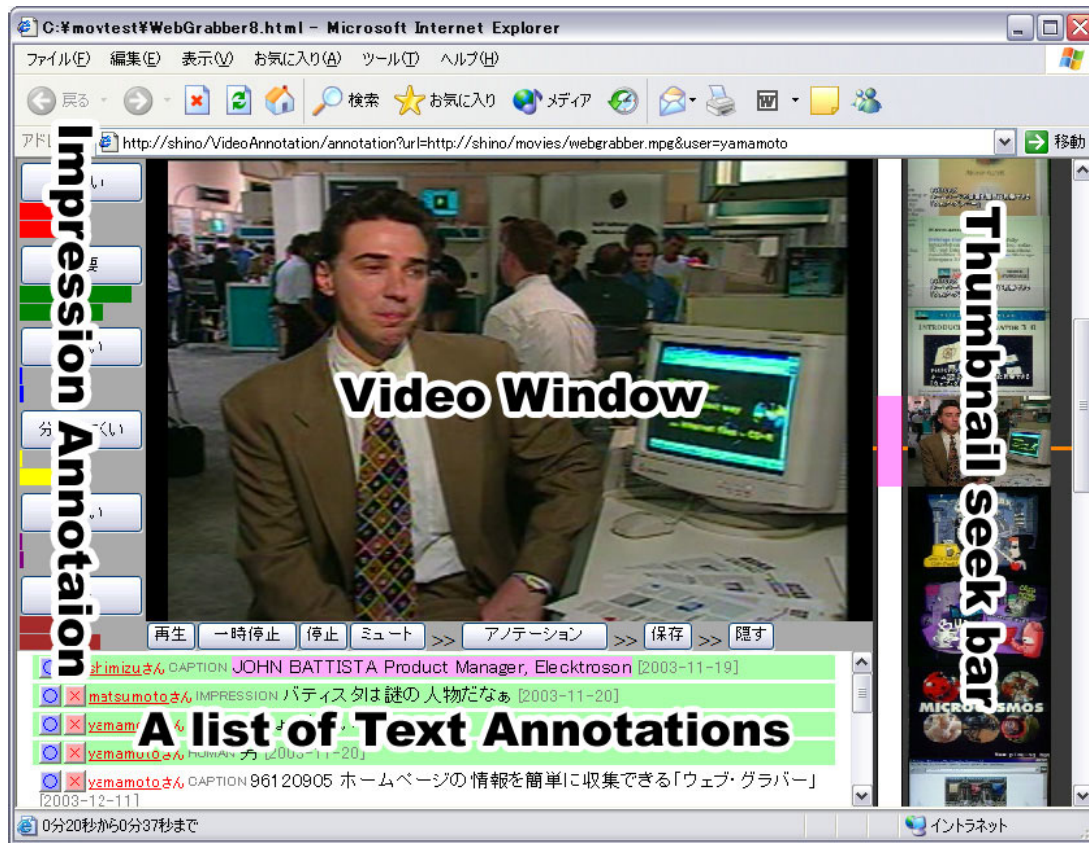
2. Impression-Annotation:

Bei diesem Prozess kann jedes Video mit verschieden vorgegebenen Eindrücken durch einen Mausklick verbunden werden. Die Anzahl der angeklickten Beschreibungen sagt etwas über die Impressionsstärke des Materials aus.

3. Bewertungsannotation:

Dieses System ist zugänglich für jeden Internetnutzer. Die Annotationen müssen nicht immer richtig sein, deshalb müssen die Annotationen selektiert werden. Die Selektion basiert auf folgenden Prinzip: „Informationen, die von einer vertrauenswürdigen Person stammen, sind auch vertrauenswürdig“. Das System berücksichtigt somit keine Annotationen, die nicht glaubwürdig sind.

Abb. 5. Graphische Oberfläche des iVAS Tool



Quelle: [11]

5. Soccer highlight automatic video annotation [12]

Bei dem System handelt es sich um ein automatisiertes Annotationssystem, was bedeutet, dass alle Beschreibungen (Metadaten) von dem System erstellt werden.

Die Semantische Annotation von Fussballvideos wird durchgeführt, indem Angriffe und gezielte Schüsse erkannt werden. Die Aktionen im Video werden durch die Verwendung von verschiedenen Merkmalen, die vom Video extrahiert werden, bestimmt:

- Bewegungsvektoren
- YUV Farbkomponenten
- Verhältnis zwischen den Pixel der Spieler der beiden Teams (die Farben beider Teams sind a-priori bekannt)
- Spielfeldlinien

Folgende Aktionen kann dieses System erkennen:

- Schuss auf das Tor – SOG (shot on goal)
- Angriff – AA (attack action)
- gezielter Schuss – PK (placed kick)
- Aktion, die ohne Schuss aufs Tor endet – NSOG (no shot on goal)
- andere Aktionen – OA (other action)

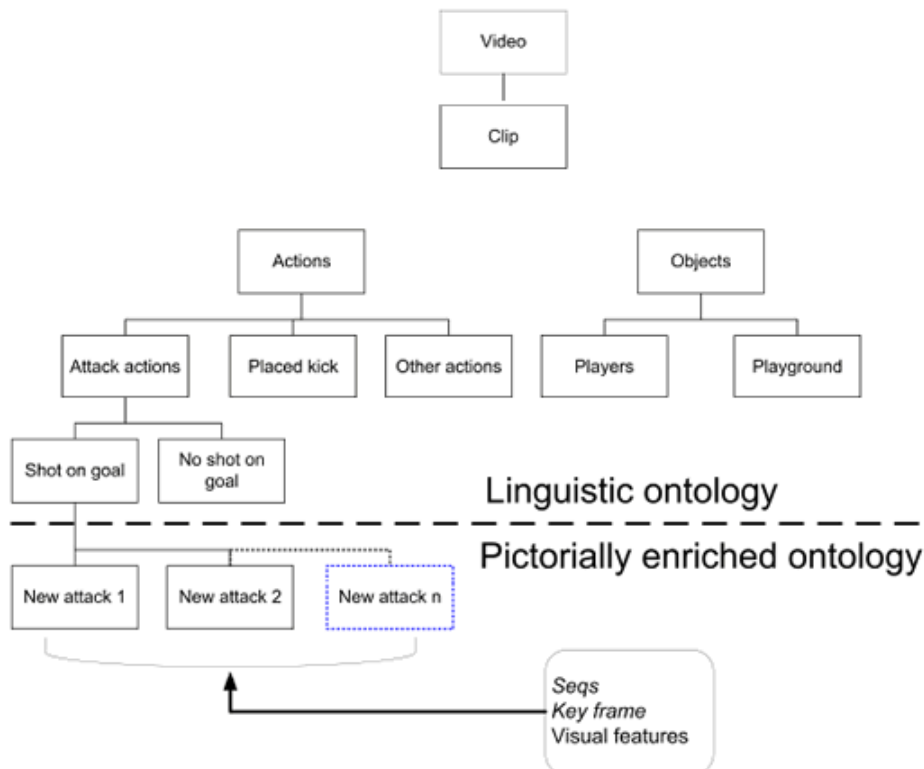
Abb. 6. Test Ergebnisse von Soccer highlight automatic video annotation

Highlight	Precision	Recall
AA	0.98	0.88
PK	0.63	0.91
SOG	0.96	0.88
NSOG	0.74	0.80
OA	0.77	0.95

Quelle: [12]

Aktionen, Spieler und Spielfeld, die erkannt wurden, werden mit textuellen Ontologien verbunden. Dieses System verwendet durch Bilder erweiterte Ontologien. Die Idee besteht darin, dass jede Aktion, z.B. Schuss aufs Tor, viele Subklassen hat, die schwer durch einen Text zu beschreiben sind. Durch Bilder erweiterte Ontologien werden eingesetzt, um Subklassen von Aktionen zu definieren. In diesem Fall sind die Ontologien visuelle Konzepte, die vom Video extrahiert werden und verschiedene Formen haben: Sequenzen, Keyframe, Region, visuelle Merkmale (Trajektorien, Bewegungsfelder...). Sie gelten als Prototypen von verschiedenen Subklassen von Ereignissen (Aktionen) und werden als Zentrum von Clusters eingesetzt. Durch die Verwendung von „Fuzzy-c-means Clustering“-Algorithmen [13] werden die Clips in Subklassen gruppiert und das Zentrum von jedem Cluster ist eine visuelle Ontologie.

Abb. 7. Durch Bilder erweiterte Ontologien



Quelle: [12]

4. Schlusswort

Metadaten bieten eine strukturierte Beschreibung der Daten, die dem Computer erlauben, auf der Basis von Bedeutung zu operieren. Auswertungen visueller Szenen auf semantischer Ebene stellen hohe Anforderungen an die Erstellung von Multimedia-Metadaten. Die erfassten Metadaten sollen den Umgang mit Multimedia-Dokumenten verbessern. Hauptsächlich soll das Auffinden und Beschaffen dieser Dokumente vereinfacht werden.

Für viele Annotationssysteme ist es heutzutage kein Problem mehr, Low-Level-Eigenschaften aus Multimedia-Daten zu extrahieren aber das Herausfiltern von High-Level-Eigenschaften ist immer noch schwierig. Dabei spielen die High-Level-Eigenschaften für die Entwicklung effizienter Multimedia-Anwendungen ebenso eine große Rolle [7].

Es ist jedoch auch möglich Systeme zu entwickeln, die durch Low-Level-Beschreibungen (also Größe, Form, Oberflächenbeschaffenheit) auf die Bedeutung (High-Level-Szenen-Beschreibungen) schließen. Man spricht von vollautomatischen Annotationssystemen. Diese Systeme sind hauptsächlich für sehr spezifische Lebensbereiche entwickelt worden, wie z.B. Medizin oder Sport. Wichtig ist es, dass die Systeme mit den schon vorhandenen Regeln der jeweiligen Bereiche gefüttert werden. Dadurch ist es dem System möglich eigene Schlussfolgerungen zu ziehen, wie z.B. das Überqueren der Torlinie des Balles als Tor zu werten. Es ist jedoch schwer ein vollautomatisches Annotationssystem zu entwickeln, das ohne Begrenzungen für jeden Bereich anwendbar ist.

Literaturverzeichnis

- [1] Robert Strötgen, Universität Hildesheim, *Metadaten-Anwendungen und „Dublin Core“*, 30.04.2003, zuletzt besucht: 12.06.2006.
- [2] Wikipedia: *Interoperabilität*, http://de.wikipedia.org/wiki/Interoperabilität_, zuletzt besucht: 12.06.2006.
- [3] S.Brenner, T.Gaab, T.Klöpfel, M.Naujoks, K.Schuchhardt, N.Sinemus, M.Stelter, M. Düro, *Metadaten nach dem Dublin Core Metadata Element Set in ausgewählten bibliothekarischen Projekten*, 21.04.2005, <http://publikationen.ub.uni-frankfurt.de/volltexte/2005/750/>, zuletzt besucht: 12.06.2006.
- [4] Matthias Glatschke, *Im Musikhandel wird es immer wichtiger gefunden zu werden*, <http://www.semantic-web.at>, zuletzt besucht: 13.06.2006
- [5] Wikipedia, *Exchangeable Image File Format*, http://de.wikipedia.org/wiki/Exif_, zuletzt besucht: 12.06.2006.
- [6] W3C, World Wide Web Consortium *Image annotation on the Semantic Web*, http://www.w3.org/2001/sw/BestPractices/MM/image_annotation.html, zuletzt besucht: 12.06.2006
- [7] Daniel Bortey, Technische Universität Hamburg, *Die Verwendung des MPEG-7 Standards in der Szenen Interpretation*, 15.10.2005, <http://www.sts.tu-harburg.de/pw-and-m-theses/2005/bort05.pdf>.
- [8] IBM Research, *VideoAnnEx Annotation Tool*, <http://www.research.ibm.com/VideoAnnEx/>, ausgewertet am 13.06.2006
- [9] Ching-Yung Lin, Belle L. Tseng, Milind Naphade, Apostol Natsev and John R. Smith IBM T. J. Watson Research Center, *MPEG-7 Video Automatic Labeling System*
- [10] R.Goulate, R.G. Cattelan, J.A.Camacho-Guerrero, V.RInacio Jr., M.C.Pimentel, University of Sao Paulo, *Interactive Multimedia Annotations: Enriching and Extending Content*
- [11] D.Yamamoto, K.Nagao, Nagoya University, *iVAS: Web-based Video Annotation System and its Applications*
- [12] M.Bertini, R.Cucchiara, A.Del Bimbo, C.Torniai, Università di Firenze, Università di Modena e Reggio Emilia, *Ontologies Enriched with Visual Information for Video Annotation*
- [13] J.C.Bezdek, *Pattern Recognition with Fuzzy Objective Function Algorithms*, Plenum Press, New York, 1981.