

UNIVERSITÄT MANNHEIM

KOLLABORATIVE ANNOTATIONSSYSTEME FÜR VIDEOS

Seminararbeit

eingereicht im: Juni 2006

von: Thomas Plotkowiak
geboren am 09. Juli 1981
in Zabrze

Matrikelnummer: 0933679

Universität Mannheim
Lehrstuhl für Praktische Informatik 4
D – 68159 Mannheim
Telefon: +49 0621 181-2600, Fax +49 0621 181-2601
Internet: <http://www.informatik.uni-mannheim.de/pi4/>

Inhaltsverzeichnis

Stichwortverzeichnis	iv
Abbildungsverzeichnis	v
Abkürzungsverzeichnis	vi
1 Einleitung	1
1.1 Aufbau der Arbeit	1
2 Umfeld und Geschichte der Video Annotation	3
2.1 Audiovisuelle Bibliotheken	3
2.2 Die Annotierung und Beschreibung von audiovisuellen Bibliotheken	3
2.3 Die kollaborative Video Annotation	4
2.4 Die TREC-Konferenz	4
2.5 Die TRECVID-Konferenz	5
3 Bereits bestehende Video-Annotations Tools	6
3.1 Kollaborative asynchrone verteilte Video Annotations Tools . . .	7
3.2 Kollaborative synchrone verteilte Video Annotations Tools . . .	7
3.3 Asynchrone Systeme vs. Synchrone Systeme	8
4 IBM Research - VideoAnnEx 2.0	9
4.1 Übersicht über die Methodik des Systems	9
4.2 Übersicht über die Komponenten des Systems	9
4.3 Fazit	11
5 IBM Research - Efficient Video Annotation System (EVA)	12
5.1 Ziele des „Efficient Video Annotation Systems“	12
5.2 Nötige Annahmen	12
5.3 Übersicht über die Komponenten des Systems	12
5.4 Übersicht über die Methodik des Systems	13
5.5 Fazit	14
6 University of Queensland - Vannotea (FilmEd Projekt)	15
6.1 Ziele des Vannotea-Systems	15
6.2 Übersicht über die Architektur des Systems	15
6.3 Übersicht über die Komponenten des Systems	15
6.4 Benutzerschnittstelle des Systems	17
6.5 Fazit	18

7 University of Indiana - eSports System	19
7.1 Ziele des eSports Systems(Gang Zhai u. Bulut 2004)	19
7.2 Übersicht über die Architektur des Systems	19
7.3 Übersicht über die Methodik des Systems	20
7.4 Übersicht über die Komponenten des Systems	20
7.5 Fazit	21
Literaturverzeichnis	22

Stichwortverzeichnis

Kollaborative Annotationssysteme

Video Annotation

Vannotea

VideoAnnEx

eSports

Narada Brokering

Grange Net

Efficient Video Annotation

TRECVID

Abbildungsverzeichnis

4.1	Benutzerschnittstelle des VideoAnnEx Systems	10
5.1	Benutzerschnittstelle des EVA Systems	13
6.1	Schematische Darstellung des .NET Remoting Event Handling .	16
6.2	Benutzerschnittstelle des Vannotea Systems	17
7.1	Schematische Darstellung der Architektur des eSports Systems .	20

Abkürzungsverzeichnis

TRECVID	TREC Video Retrieval Evaluation
TREC	Text Retrieval Conference
WWW	World Wide Web
RTSP	Real-Time Streaming Protocol
FF	Fast Forward
FR	Fast Rewind
EVA	Effective Video Annotation
MMCS	Multimedia Collaboration System
ZGDV	Zentrum für Graphische Datenverarbeitung
ID	Identification
MPEG	Motion Picture Expert Group
XGSP	XML based General Session Protocol

1 Einleitung

Obwohl Indizierungs- und Annotations-Programme für Textdokumente, Webseiten, Bilder, Audio- und Videoressourcen schon existieren, wurden die entsprechenden Annotationsprogramme für Einzelanwender konzipiert. Typischerweise können die Beschreibungen und Annotationen zusammen genutzt werden, indem sie auf einem Server gespeichert werden. Aber die Annotations-Applikationen an sich wurden nicht für eine gemeinsame Arbeit am Video- und Audio-Material entworfen.

Die Systeme, die in dieser Seminararbeit beschrieben sind, behandeln sowohl die asynchrone, kollaborative Annotation als auch die synchrone Echtzeit-Annotation, sowie auch zugehörige kollaborative Diskussion über Videomaterial.

1.1 Aufbau der Arbeit

Die Seminararbeit ist wie folgt strukturiert: Kapitel 2 verschafft eine Motivation, warum kollaborative Annotationssysteme gebraucht und entwickelt werden. Gleichzeitig werden die verschiedenen Dimensionen der Annotation erläutert. Es wird eine Übersicht über das Umfeld in dem diese Applikationen benutzt werden verschafft: Dazu gehören eine Übersicht über die TREC-Konferenz, sowie auch eine Übersicht über die TRECVID-Konferenz.

Kapitel 3 stellt die verschiedenen, bisher vorhandenen Einzelanwender-Video-Annotationssysteme vor und grenzt diese von den kollaborativen Annotationssystemen ab. Es folgt eine Übersicht der vorhandenen kollaborativen Systeme. Das Kapitel zeigt Einsatzbereiche und Anforderungen solcher Systeme. Ebenfalls wird eine Abgrenzung von Post-Annotationssystemen gegenüber Echtzeit-Annotationssystemen vorgenommen.

Die Kapitel 4 und 5 behandeln die asynchronen Post-Annotationssysteme, dh. Systeme die entworfen wurden, um bestehendes Videomaterial mit Annotationen zu versehen, wobei der Hauptaugenmerk dieser Systeme die Verwendung in den oben genannten TRECVID-Konferenzen darstellt.

In Kapitel 4 wird das IBM Research Tool VideoAnnEx 2.0 vorgestellt. Es werden die Grundzüge der Architektur vorgestellt, sowie auch ein Überblick über das Grafische Benutzer Interface gegeben. Im Detail werden der Ontologie-Editor und die Gruppenadministrationsmethodik vorgestellt. Im Abschluss werden die Vor- und Nachteile dieses Systems vorgestellt und die Motivation erörtert, warum ein effizienteres System entwickelt worden ist.

In Kapitel 5 wird das EVA System als Nachfolger des VideoAnnEx 2.0 vorgestellt. Es wird eine Übersicht gegeben welche Annahmen getroffen wurden,

um sich vom alten System abzugrenzen und es wird ein kurzer Überblick über die Funktionsweise und Architektur dieses Systems gegeben. Im Nachfolgenden wird die Annotationsmethodik, sowie die Oberfläche des Programms vorgestellt. Eingehend diskutiert wird der optimierte Annotationsprozess, sowie die interne Statistikfunktion. Den Abschluss bilden die Evaluation des des Systems sowie ein Fazit mit den Vor- und Nachteilen.

Die Kapitel 6 und 7 behandeln die synchronen Echtzeitannotations-Systeme, die in erster Linie dazu entworfen wurden, um Annotationen kollaborativ zu tätigen und diese in einer breiten Gruppe von verteilten Mitgliedern auszutauschen. Diese Systeme finden vielseitige Anwendungen, auf die in den Kapiteln einzeln eingegangen wird.

Kapitel 6 stellt das Vannotea (FilmEd) System vor. Es wird eine Motivation gegeben, warum synchrone Echtzeit-Systeme benötigt werden. Entsprechend werden die verschiedenen Anwendungsbereiche des Vannotea Systems vorgestellt. Im Detail wird die grafische Benutzeroberfläche, sowie die benötigte Architektur vorgestellt. Es wird dabei die .NET Remote Architektur vorgestellt und eine Abgrenzung zu vorhandenen Lösungen vorgenommen. Auch wird ein neuer Ansatz vorgestellt, wie subjektive und objektive Annotationen getrennt zu speichern sind. Im Abschluss wird ein Fazit über die Vor- und Nachteile des Systems gebildet.

In Kapitel 7 wird das eSports System präsentiert, wobei im Detail zwei besondere Architekturmerkmale des Systems hervorgehoben werden (NaradaBrokering, GlobalMMCS). Es wird ebenfalls die Funktion und Methodik des Systems erläutert, wobei der Hauptaugenmerk auf der praktischen Anwendung des Systems im Sportbereich gezeigt werden soll.

2 Umfeld und Geschichte der Video Annotation

Das Umfeld, in dem die asynchronen, kollaborativen Video Annotations-Applikationen benutzt werden, ist bisher sehr begrenzt. Um dieses Umfeld vorzustellen, wird eine Übersicht über sowohl über die TREC-Konferenz als auch über die TRECVID-Konferenz geben. Die Einleitung über die Geschichte dieser Systeme und der Nutzen solcher audiovisueller Bibliotheken gehen voraus.

2.1 Audiovisuelle Bibliotheken

Über Jahre hinweg mussten Akademiker in Bibliotheken gehen um Wissen, das meist in Buchform enthalten war, abzurufen. Bibliotheken organisierten die Bücher nach Interessensgebieten, die für die entsprechenden Gruppen am zutreffendsten waren. Später wurden Datenbanksysteme eingeführt, die eine mächtigere Suche auf dem Bestand erlaubten. Schließlich mit der Erfindung des World Wide Web wurden Suchmaschinen und digitale Bibliotheken erfunden. Diese erlaubten Menschen auf der ganzen Welt, Wissen, das digital gespeichert wurde, überall zu publizieren, miteinander zu teilen, danach zu suchen und es abzurufen. Mit der Zeit wurden nicht nur Dokumente in Textform, sondern auch Dokumente in audiovisueller Form mehr und mehr verbreitet. Dies lag nicht zuletzt daran, dass die entsprechenden Bandbreiten im Internet und die Datenspeicher vor Ort viel größere Mengen an Datenvolumen erlaubten.

Audiovisuelle Bibliotheken werden ihren Platz in verschiedenen Gebieten finden, wie der Bildung, der Wissenschaft und der Videoproduktion oder dem Entertainment. Im Bildungssektor versuchen schon heutzutage viele Universitäten die aufgezeichneten Vorlesungen online verfügbar zu machen. Auch in der Wissenschaft werden große Datenmengen an Video- und Filmmaterial erzeugt, wie z.B. in der Medizin. Um dieser „Datenflut“ gerecht zu werden, braucht man neue Technologien, um diese Daten wieder aufzufinden und verfügbar zu machen.

Um diesen Prozess effektiv zu gestalten und um komplexe Anfragen zu stellen, brauchen wir eine Indizierung des gesamten Bestandes. Diese kann entweder manuell oder automatisch stattfinden. Diese „Metadaten“ müssen erfolgreich mit dem bestehendem Material verknüpft werden, was schon heute mit Hilfe von z.B. MPEG7-Deskriptoren geschieht.

2.2 Die Annotierung und Beschreibung von audiovisuellen Bibliotheken

Unter dem Begriff der Beschreibung versteht man die objektive, detaillierte High-level-Beschreibung des Materials, die meistens nur durch einen professio-

nellen Anwender durchgeführt werden kann. Hier wird der Kontext des Materials semantisch erfasst und beschrieben. Unter der Annotierung versteht man die unstrukturierte Anreicherung des Materials mit automatisch generierten Annotationen, die bereits automatisch errechnet werden können. Unter der Annotierung versteht man aber auch die subjektive Beschreibung des Materials durch einen Menschen, der seine persönliche Meinung festhalten möchte.

Video-Annotationssysteme sollten beide Aspekte beherrschen, oft findet jedoch eine scharfe Trennung zwischen diesen beiden Begrifflichkeiten nicht statt. Es hat sich herausgestellt, dass diese Unterscheidung ein wichtiges Thema ist und maßgeblich zur Effektivität solcher Systeme beiträgt.

2.3 Die kollaborative Video Annotation

Ein System, das die Suche nach audiovisueller Information und nach persönlichen Annotationen erlaubt, sollte auch zulassen seine Annotationen untereinander mit anderen Menschen zu teilen, oder auch Fragen und Erläuterungen untereinander zuzulassen. Die kollaborative Video-Annotation ist gleichzeitig auch ein Weg das sogenannte „Gruppenwissen“ abzuspeichern.

Es gibt mehrere Grade der Kollaboration: Einerseits lässt sie sich nach zeitlichen Aspekten gliedern, womit die synchrone und asynchrone Annotation gemeint ist. Gleichzeitig ist auch eine weitere Dimension vorhanden, in der man eine räumliche Gliederung vornimmt: Die sogenannte „Face-to-Face“ Annotation und die verteilte Annotation.

In dieser Seminararbeit behandle ich beide zeitlichen Aspekte und stelle jeweils zwei Systeme vor, die für die synchrone und asynchrone Annotation entwickelt wurden.

2.4 Die TREC-Konferenz

Einen sehr wichtigen Beitrag zur Entwicklung kollaborativer Systeme hat die TREC-Konferenz (Voorhees 2006) geleistet. Die TREC oder auch „Text REtrieval Conference“, wird vom National Institute of Standards and Technology (NIST 2006) und dem U.S.-Verteidigungsministerium finanziert. Sie wurde im Jahre 1992 ins Leben gerufen, mit der Absicht eine Plattform zu schaffen, die die Forschung im Bereich der Informations-Indizierung vereint.

Neben der TREC-Konferenz existieren auch ähnliche Veranstaltungen, wie die CLEF (Cross-Language Evaluation Forum), (CLEF 2006), die sich hauptsächlich mit der semantischen Analyse von nicht englischsprachigen Dokumenten befasst. Ebenso existieren Konferenzen im asiatischen Raum, die sich spezifisch mit dem „Information Retrieval“ in den Landessprachen wie Japanisch oder Chinesisch befassen (Chinese Web test collection, NTCIR Workshops), (CWIRF 2006).

Im Laufe der Zeit wurden im Rahmen der TREC weitere Medien zur Informations-Indizierung hinzugefügt. So wurden im Jahre 2001 zum ersten Mal Videos untersucht, um die Forschung auf dem Bereich der automatischen Segmentierung, Indizierung, und der inhaltsbezogenen Suche auf einer Konferenz zu vereinen. Aus dieser Initiative wurde im Jahre 2003 die TRECVID (TRECVID 2006) geboren.

2.5 Die TRECVID-Konferenz

Hervorgegangen aus der oben genannten TREC-Konferenz, ist die TRECVID mittlerweile ein unabhängiger Workshop, der sich zwei Tage lang mit Themen aus dem Bereich der Video-Informationsgewinnung befasst. Über 40 Teams aus Europa, Amerika, Asien und Australien treffen sich jährlich, um immer neue Entwicklungen auf diesem Forschungsgebiet auszutauschen. Speziell im Bereich der kollaborativen Video-Annotation ist hier die TRECVID im Jahre 2002 hervorzuheben, die sich zum ersten Mal damit befasste, ein „**ground truth based system**“ für Videos zu erzeugen, welches später als Ausgangsmaterial für z.B. maschinelles Lernen dienen sollte. Dazu wurden 68,45 Stunden MPEG1 Video Material aus verschiedenen Bereichen annotiert. Durch diese große Menge an Daten wurde zum ersten Mal der kollaborative Aspekt einer solchen Konferenz genutzt, indem das Archiv gleichzeitig von allen teilnehmenden Gruppen annotiert wurde. Hier kam zum auch ersten Mal das „IBM Research Tool VideoAnnEx“(IBM 2006) zum Einsatz, welches für die Konferenz zu einem kollaborativen System erweitert wurde.

Im Jahre 2005 wurde auf der TRECVID der gleiche Ansatz verfolgt, jedoch wurde hierbei ein anderes kollaborative Video Annotations Tool verwendet (Efficient Video Annotation) (IBM 2005), welches bereits erkannte Schwächen des Vorgängers VideoAnnEx ausbesserte und gleichzeitig bessere Ergebnisse versprach. Auch hier wurden 120 Stunden Videomaterial annotiert, das hauptsächlich aus Nachrichten-, Meeting- und Konferenz-Mitschnitten bestand.

Was die kollaborative Video-Annotation angeht, waren diese beiden Konferenzen von größtem Interesse, da hier die entsprechenden Tools im professionellen Bereich eingesetzt und evaluiert wurden.

Neben der TRECVID existieren auch ähnliche Workshops, die sich mit dem gleichen Themengebiet befassen, dazu gehören die IEEE International Conference on Multimedia and Expo (ICME), International Conference on Image and Video Retrieval (CIVR) , International Workshop on Adaptive Multimedia Retrieval sowie die International Conference on Multimedia Modeling (MMM 2007).

3 Bereits bestehende Video-Annotations Tools

Folgende Systeme sind aus meiner persönlichen Sicht in der Einzelanwender-Video-Annotation am meisten vorangeschritten, werden jedoch mangels ihrer kollaborativen Funktionalität nicht in dieser Arbeit besprochen:

- Virage - Video Logger
- IBM MPEG7 Annotation Tool
- Ricoh - Movie Tool
- ZGDV - VIDETO
- COALA - LogCreator

Der VideoLogger von Virage ist eine professionelle Anwendung, die in der TV Produktion verwendet wird. Sie kann Szenen extrahieren, Überschriften erkennen, und Videotextinformationen auslesen, um einen strukturierten Index zu erzeugen. Außerdem existieren Plugins, um eine Spracherkennung auszuführen. Die Annotationen werden leider nicht MPEG-7 konform abgespeichert und es werden auch keine offenen Standards, wie DublinCore(DUBLINCORE 2006), unterstützt. Ebenso unterstützt VideoLogger keine hierarchische Segmentierung, Regionale Deskriptoren, und auch keine kollaborative Annotierung.

IBM's MPEG 7 Annotation Tool (VideoAnnEx 1.0) unterstützt regionale Deskriptoren, Szenen Extrahierung und ein anpassbares Lexikon. Größte Nachteile sind eine fest eingestellte Videogröße und die Anforderung, jeweils drei Kategorien einer Szene gleichzeitig annotieren zu müssen. Ebenso war Anfangs keine kollaborative Annotierung möglich, die jedoch in Version 2.0 hinzugefügt wurde.

Ricoh's Movie Tool, unterstützt die hierarchische Segmentierung, ebenso lässt bei der Szenen Extrahierung Einstellungen im Schwellwertbereich zu. Das MovieTool ist zwar ein sehr ausgereiftes System, stellt jedoch zu hohe Ansprüche an den Endanwender, der sich mit einer sehr unzugänglichen Benutzerschnittstelle auseinandersetzen muss. Diese ist zwar sehr nahe am MPEG-7 Standard angelegt, erfordert jedoch genaue Kenntnisse, da man sich das komplexe XML Schema aneignen muss, um mit dem Programm zu arbeiten. Ebenfalls fehlt in diesem System die Möglichkeit Videos kollaborativ zu annotieren.

Der LogCreator des COALA Projekts, ist ein Internet basiertes Tool, das Video-Annotationen zulässt. Es verfügt über eine automatische Szenen Extrahierung und eine gute Benutzerschnittstelle, um eine hierarchische Segmentierung vorzunehmen. Trotzdem ist das Programm nur sehr eingeschränkt nutzbar, da es speziell für den Bereich der Fernsehnachrichten Annotierung entwickelt wurde und einer fest vorgegebenen Struktur folgt. Ebenfalls fehlt ein kollaborativer Modus.

3.1 Kollaborative asynchrone verteilte Video Annotations Tools

Folgende Systeme sind aus meiner persönlichen Sicht in der kollaborativen Video-Annotation am meisten vorangeschritten:

- IBM MPEG7 Annotation Tool (VideoAnEx 2.0)
- IBM MPEG7 Efficient Video Annotation Tool (EVA)

IBM's MPEG7 Annotation Tool (VideoAnEx 2.0) unterstützt regionale Deskriptoren, Szenen Extrahierung und ein anpassbares Lexikon. Größte Nachteile sind eine fest eingestellte Videogröße und die Anforderung, jeweils drei Kategorien einer Szene gleichzeitig annotieren zu müssen. Die kollaborative Video-Annotierung ist durch ein Webinterface möglich. Es können bestimmten Benutzern Teile des zu annotierenden Materials zugeteilt werden.

IBM MPEG7 Efficient Video Annotation Tool (EVA) arbeitet webbasiert und bietet somit eine gute und kostengünstige Möglichkeit, das System kollaborativ zu benutzen. Die Szenen werden im Voraus extrahiert und der Benutzer mit schon geschnittenem Material versorgt. Es bietet eine sehr ausgereifte Benutzeroberfläche, die ein effizientes und schnelles Annotieren ermöglicht. Außerdem verfügt es über eine statistische Auswertung der Annotationen, was bisher auf diesem Gebiet einzigartig ist.

3.2 Kollaborative synchrone verteilte Video Annotations Tools

- FilmEd - Vannotea
- eSports System

Vannotea, hervorgegangen aus dem FilmEd Projekt der Universität Queensland, bietet eine Szenen-Extrahierung, eine Unterstützung für MPEG1,2 und 4 Formate, einen Dublincore und MPEG7 Deskriptorenstandard, und speziell in .NET implementierte Protokolle, die ein synchrones Arbeiten mit hoher Bildwiederholfrequenz erlauben.

Das eSports System der Universität Indiana bietet keine Szenen-Extrahierung, erlaubt dafür durch eine Komponenten basierte Architektur, eine Erweiterung des Systems über Plugins, Unterstützung für MPEG1 Formate und basiert auf einer speziellen für die Echtzeit-Umgebung geschaffenen Architektur (GlobalMMCS, NaradaBrokering). Statt einer MPEG7 konformen Annotierung bedient es sich des „Media Enrichments“, welches später beschrieben wird.

3.3 Asynchrone Systeme vs. Synchrone Systeme

Wie bereits erläutert, existieren beide Systeme nebeneinander und stehen nicht in direkter Konkurrenz. Dies lässt vermuten, dass die synchronen, kollaborativen Systeme ein anderes Einsatzgebiet besitzen als die asynchronen Systeme.

Während asynchrone Systeme dazu benutzt werden, um bestehende Archive möglichst umfassend zu kategorisieren und zu annotieren, werden synchrone Systeme zu einer kollaborativen Echtzeit-Kommunikation und Bearbeitung von sehr aktuellen, wenn nicht sogar live eingespieltem Material benutzt.

Bei asynchronen Systemen liegt das Hauptaugenmerk auf der Detailgenauigkeit und der feinen Granularität der Annotationen, die möglichst objektiv und umfassend die einzelnen Szenen beschreiben sollen. Hierbei ist die Qualität der Annotationen von besonders hohem Interesse, da diese Daten später für weitere Projekte verwendet werden und keinem eigentlichen Selbstzweck dienen. Mögliche Einsatzgebiete sind hier die künstliche Intelligenz mit z.B. dem „machine learning“.

Bei synchronen Systemen liegt das Hauptaugenmerk auf dem möglichst schnellen und unproblematischen Zugriff auf das Medium und die entsprechenden Annotationen. Es wird hier Aspekte wie z.B. ein möglichst unkompliziertes und effizientes Teamwork geachtet, wobei die Teilnehmer einer solchen Konferenz ihr Wissen live mit anderen, entfernten Mitarbeitern austauschen. Der Grad der Annotationen ist hierbei von höchstem Maße subjektiv. Es wird nicht unbedingt genau das annotiert, was unmittelbar zu sehen ist, sondern was in einem gegebenen Kontext von Interesse für die Teilnehmer ist. Der wichtigste Aspekt ist, dass sämtliche Benutzereingaben sofort auch für die anderen Teilnehmer sichtbar sind, um einen möglichst hohen Grad an Kommunikation und Interaktivität zu gewährleisten.

4 IBM Research - VideoAnnEx 2.0

4.1 Übersicht über die Methodik des Systems

VideoAnnEx v2.0 (Ching-Yung Lin u. Smith 2002) erlaubt eine kollaborative Annotation zwischen verschiedenen Benutzern über das Internet. Die Benutzer selbst loggen sich mit ihrer Benutzeridentifikation und Passwort auf einem zentralen Server ein. Dieser speichert zentral alle MPEG7-Dateien, verwaltet die kollaborative Arbeit und überwacht den Annotationsprozess.

Für eine kollaborative Annotation gibt es drei Rollen die ein Benutzer einnehmen kann:

- Projekt Manager
- Gruppen-Administrator
- genereller Benutzer

Der **Projekt Manager** administriert den VideoAnnEx Server, koordiniert die 'IDs' und teilt den Gruppen ihren Teil an zu annotierbaren Medien zu.

Der **Gruppen-Administrator** koordiniert den Annotationsprozess indem er die Mitgliedern der Gruppe mit Teilen der zu annotierbaren Medien versorgt.

Der generelle **Benutzer** annotiert das Videomaterial.

4.2 Übersicht über die Komponenten des Systems

Die Segmentierungs-Komponente teilt das Video in kleinere Teile, die dann entsprechend von den Gruppen Administratoren an ihre Mitglieder weitergeleitet werden.

Das semantische Lexikon definiert die auswählbare Annotations-Domäne. Hierbei lädt sich ein Benutzer immer gleichzeitig einen Teil des Videos mit entsprechendem Lexikon vom Server herunter.

Ein automatischer Annotations-Ergänzung-Mechanismus, erleichtert den Benutzern den Prozess, in dem er bei ähnlichen Szenen auch ähnliche Begriffe vorschlägt.

Die grafische Benutzerschnittstelle (zu sehen in Abb. 4.1) ist in vier Regionen unterteilt, von denen jede einer speziellen Aufgabe dient. Das „Video Playback“ Fenster versorgt den Benutzer mit der benötigten visuellen Information. Das „**Shot Annotation**“ Modul zeigt die annotierten Begriffe zusammen mit

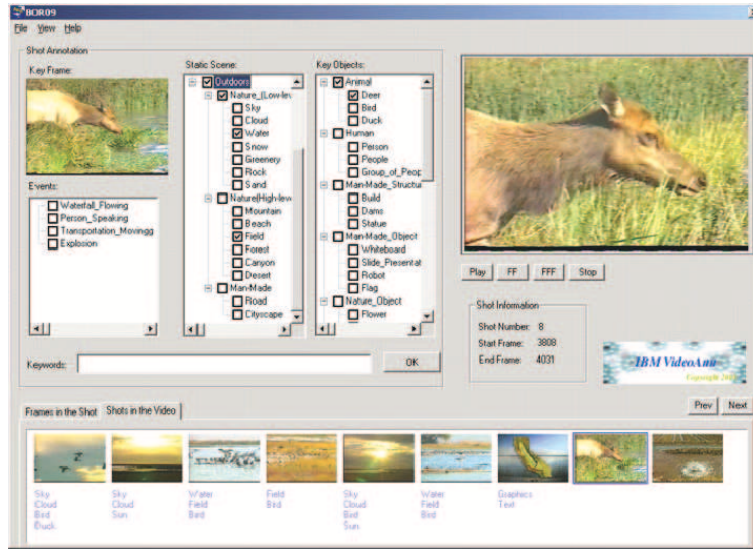


Abbildung 4.1: Benutzerschnittstelle des VideoAnnEx Systems

dem zugehörigen „Keyframe“ an. Hierbei können auch die einzelnen Bilder, die zu einer Szene gehören, durchgesehen werden. Das „**Region Annotation**“ Modul erlaubt es dem Benutzer Regionen auf einem interessantem „Frame“ einzuzichnen, um diese speziell zu annotieren. Die Szenen-Segmentierung an sich beruht auf der Differenz in den Farbhistogrammen und den Kantenwerten.

Nachdem das Material in Szenen segmentiert wurde, trägt der Benutzer im Ontologie-Editor die Beschreibung dieses ein. Hierbei kann er aus 139 fest definierten Stichpunkten wählen, die sich in drei Kategorien aufteilen: „**Static Scene**“ - die Umgebung, „**Key Object**“ - das Hauptmotiv und „**Event**“ - den aktiven Vorgang an sich. Kritisch hierbei ist, dass der Benutzer alle drei Aspekte auf einmal notieren muss. Dies führt zu einer ineffektiven Arbeitsweise, was bei einer späteren Evaluierung des Systems festgestellt wurde.

Ein besonderes Merkmal ist das „**Template Matching**“: Hierbei kann der Benutzer einen spezifischen Bereich auswählen, für den Szenen gefunden werden, die eine hohe Ähnlichkeit aufweisen. Hierbei wird auf fünf aufeinander folgende Frames ein Median Filter gelegt und nach einem Schwellwertverfahren in Bezug auf Farb- und Kanteneigenschaften entschieden, welche Szenen sich der Auswahl besonders ähneln:

$$S = \delta(S_C > \tau'_C) \delta(S_E > \tau'_E)$$

$$S_C = \frac{1}{N} \sum \delta(d(P_C, P_{MC}) > \tau_C)$$

$$S_E = \frac{1}{N} \sum \delta(d(P_E, P_{ME}) > \tau_E)$$

C repräsentiert die Farbeigenschaften und E die Kanteneigenschaften. $\tau_C, \tau_E, \tau'_C, \tau'_E$, sind die entsprechenden Schwellwerte. $\delta()$ ist die binäre Entscheidungsfunktion, und $d()$ ist die euklidische Distanz der beiden Testregionen im Raum. N ist die Anzahl der Pixel in der Region.

4.3 Fazit

Das System hatte sich auf der TRECVID 2003, seinem Debüt, durchaus bewährt. Nichts desto trotz gab es Punkte die von den Teilnehmern in einer späteren Evaluierung als verbesserungswürdig bezeichnet wurden. Auf der Wunschliste standen:

- Eine bessere Hilfefunktion und Richtlinien, wie eine vorschriftsmäßige Annotation aussehen soll.
- Das Heraussuchen des richtigen Begriffs im Lexikon sollte beschleunigt werden. Das Scrollen im Lexikon verlangsamte den Annotationsprozess.
- Eine bessere Wiedergabe des Videos - gewünscht wurden ein „FF“ und „FR“ Mode.
- Die Möglichkeit mehrere Szenen auf einmal zu annotieren.
- Eine Vorauswahl des Genre, um die Annotierung zu beschleunigen.

Der Hauptkritikpunkt war, dass 3 verschiedene Konzepte auf einmal annotiert werden müssen. Dies verbraucht sehr viel Zeit und trägt nicht dazu bei, die Qualität der Annotationen zu verbessern. Deshalb wurde dieser Kritikpunkt und viele der oben genannten Punkte in der Nachfolgeversion (v3.0) beherzigt, was ihr auch ihren treffenden Namen gab: „Efficient Video Annotation System“ (TimoVolkmer u. Apostol Paul) . Diese wird auf den folgenden Seiten näher vorgestellt.

5 IBM Research - Efficient Video Annotation System (EVA)

5.1 Ziele des „Efficient Video Annotation Systems“

Obwohl die Vorgängerversion dieses Systems durchaus mächtig in ihrer Funktionalität war, hatte sie sich, wie zuvor erwähnt, als nicht optimal für die TRECVID-Konferenzen erwiesen.

Als Nachfolgeversion des VideoAnnEx v2.0 Systems war es Ziel des EVA Systems (Timo Volkmer u. Apostol Paul) die Bedienungsfreundlichkeit zu verbessern und die Vollständigkeit der Annotationen anzustreben. Weitere Verbesserungskriterien waren die Vereinfachung und Beschleunigung des Annotationsprozesses, ohne den Organisationsaufwand zu erhöhen. Dazu wurden spezielle Techniken entwickelt, die im Näheren vorgestellt werden.

5.2 Nötige Annahmen

Um die oben genannten Ziele zu erreichen, mussten bei diesem System folgende Annahmen getroffen werden:

- Alle Annotationen beruhen auf einem kleinen ausgewähltem Lexikon von 39 Begriffen
- Freie Annotationen sind nicht möglich
- Es gibt nur eine statische Anzeige der Szenen. Ein Abspielen ist nicht möglich.
- Alle Annotationen werden für die ganze Szene annotiert.
- Regionale Annotationen sind nicht möglich.

5.3 Übersicht über die Komponenten des Systems

Im Gegensatz zur Vorgängerversion, die noch über eine integrierte Schnitterkennung verfügte, wurde hier auf eine solche verzichtet. Das Videomaterial wird bei diesem System schon im Vorfeld geschnitten, um den Annotationsprozess zu beschleunigen und die Effizienz zu steigern.

Das Wesentliche stellt bei dem Efficient Video Annotation System die Benutzerinterface (zu sehen in Abb. 5.1) dar, die verschiedene Aspekte abdecken soll:



Abbildung 5.1: Benutzerschnittstelle des EVA Systems

- Die Möglichkeit Szenen besonders schnell ohne „Scrollvoränge“ zu annotieren.
- Den Annotationsprozess nur auf einen Aspekt zu konzentrieren.
- Bei Bedarf verschiedene Aspekte gleichzeitig zu annotieren.
- Die Möglichkeit, die Darstellungsweise der Szenen durch den Benutzer anzupassen.
- Die Möglichkeit, Maus- sowie Keyboardeingaben zu verarbeiten.

Ebenfalls wurde durch verschiedene Konfigurationsmöglichkeiten sichergestellt, dass das Benutzerinterface flexibel ist und sich auf die Vorlieben der verschiedenen Benutzer einstellen lässt.

5.4 Übersicht über die Methodik des Systems

Im vollständig webbasierten System kann der Benutzer dieses durch einen einfachen Webbrowser bedienen. Hierbei läuft ein Serverprozess auf dem Webserver, welcher sowohl Videomaterial, als auch das Lexikon zur Verfügung stellt. Beginnt der Benutzer mit der Annotation des ihm zugeteiltem Materials, werden die einzelnen Szenen in Tabellenform dargestellt. Während des Annotationsprozesses kann der Benutzer durch das ihm zugeteilte Archiv völlig frei navigieren. Die dargestellten Szenen können in vier verschiedene Kategorien klassifiziert werden:

- positive: Das Bild kann eindeutig dem gewählten Konzept zugeordnet werden.

- negative: Das Bild entspricht eindeutig nicht dem Konzept.
- ignore: Das Bild lässt sich nicht eindeutig zuordnen.
- skip: Das Bild wird übersprungen.

Ebenso trägt zur einfachen Kategorisierung der Szenen ein „Bulk“-Modus bei, in dem die oben genannten Operationen ganzen Gruppen von Szenen zugeordnet werden können. Dabei kann der Benutzer Szenen auswählen, die gleichen Inhalt besitzen und sie alle gleichzeitig mit Annotationen versehen, was den Annotationsprozess auch um ein Wesentliches beschleunigt.

Als direktes Feedback für den Benutzer existiert eine Anzeige des eigenen Fortschritts während der Arbeit, sowie eine ausgiebige Statistik über die verbrachte Zeit pro Szene, die Anzahl der annotierten Szenen und die Benutzung der Maus sowie des Keyboards. Dem Administrator des Systems stehen noch mehr statistische Möglichkeiten zur Verfügung, die den analytischen Zweck haben, die Aussagen über die Qualität der Annotationen festzuhalten.

5.5 Fazit

Durch die Reduzierung des Lexikons von 139 auf 39 Begriffe in 7 Domänen, wurde zwar einerseits ein Verlust an Detailgenauigkeit in den Annotationen bewirkt, andererseits hatte dieser durchaus positive Aspekte wie die Erhöhung der Geschwindigkeit und Effizienz der Annotationen. Ebenso erhöhten sich dadurch auch die Qualität der Annotationen, da nun zwischen den Benutzern eine hohe Übereinstimmung bei gleichen Annotationen erreicht wurde. Dies ist auch der Neuerung zu verdanken, dass Benutzer nun nicht alle Aspekte einer Szene annotieren müssen, sondern wählen können, wie viele Aspekte sie annotieren möchten. Eine Evaluierung des Systems hat gezeigt, dass, je weniger Aspekte auf einmal annotiert werden, desto höher die Übereinstimmung der Ergebnisse ausfällt und somit die Qualität dieser zunimmt. Auch die Einführung der Tastatur als Navigationshilfe und die Möglichkeit, das System ausschließlich mittels Tastatur zu bedienen, hat dazu geführt den Annotationsprozess zu beschleunigen. Durch die Einführung einer internen Statistik im System ist eine wichtige Wissensquelle geschaffen worden, die viel über die Qualität der hervorgebrachten Ergebnisse verrät.

6 University of Queensland - Vannotea (FilmEd Projekt)

6.1 Ziele des Vannotea-Systems

Im Gegensatz zu den bisher vorgestellten asynchronen Systemen, die ausschließlich für einen speziellen Zweck entwickelt wurden, ist das Vannotea System (Ronald Schroeter u. Kosovic 2004) von Grund auf als synchrones Echtzeit-Annotationssystem konzipiert worden. Ziel ist das Erzeugen einer virtuellen Konferenz, bei der die Teilnehmer MPEG2-Video in Echtzeit annotieren, synchron indizieren und auf das komplette Videoarchiv zugreifen können. Dabei sollten folgende Ziele abgedeckt werden:

- Automatische Szenenerkennung
- High Quality MPEG2-Kompatibilität
- Hierarchische Videosegmentierung
- Eine einfache Benutzerschnittstelle
- Eine hohe Domänenportabilität
- Annotationen nach MPEG7-Standard
- Granulare Annotation von Regionen, Frames und Szenen
- Objektive und Subjektive Metadaten voneinander zu trennen.

6.2 Übersicht über die Architektur des Systems

Das System besteht aus zwei voneinander getrennten Archiven für Metadaten. Das objektive Archiv ist lediglich für das Durch- und Aufsuchen des Videomaterials zuständig. Dieses wird durch speziell dafür geschulte Mitarbeiter verwaltet. Das subjektive Archiv ist für die Speicherung persönlicher Annotationen zuständig, die mit anderen Benutzern ausgetauscht werden sollen. Auf beide Archive greift der Applikations-Server zu, der als Schaltstelle zwischen den Benutzern fungiert.

6.3 Übersicht über die Komponenten des Systems

Das System besteht aus vier großen Komponenten:

- Search and Retrieval Database
- Annotation Database
- Application Server
- MPEG-2 Streaming

Die „**Search and Retrieval Database**“ ist eine Datenbank, die auf dem MPEG7-Standard aufbaut, diesen jedoch mit dem Dublin Core(DUBLINCORE 2006) Standard verbindet. Hierbei werden die Videos zuerst in Szenen zerlegt. Danach werden bibliografische Annotationen im DublinCore Standard abgespeichert, während sämtliche szenenbezogenen Annotationen regulär im MPEG7-Standard abgespeichert werden.

Die „**Annotation Database**“ speichert die Annotationen, die mit Szenen, einzelnen „Frames“ oder Regionen verknüpft werden können. Ebenfalls wird die Quelle der Annotationen gespeichert, dh. „Wer, Wann, Was“. Annotationen können hierbei subjektive Notizen, also auch Erklärungen sein. Als Basis dient der Annotea(ANNOTEA 2006) RDF-Standard, der erweitert wurde, um Annotationen auf audiovisuellem Material vornehmen zu können.

Der „**Application Server**“ ist das Herzstück dieses Systems. Bei bisherigen kollaborativen Applikationen lief normalerweise eine Master-Applikation auf einem Server, die die Maus oder Keyboardereignisse vom Client entgegennimmt und die Bildschirminformation zurückschickt (T.120, VNC Protokolle) Da man aber bei einem Echtzeitsystem für MPEG-2 viel höhere Anforderungen an die Bildwiederholfrequenz und Bandbreite stellt, sind solche Protokolle unpassend. Stattdessen wurde ein spezielles .NET Remoting (Microsoft 2006) Framework eingesetzt:

Beim **.NET Remoting Framework** (zu sehen in Abb. 6.1) können alle Sprachkonstrukte wie Konstruktoren, Schnittstellen, Methoden und Eigenschaften auf entfernte Objekte angewandt werden. Dabei verwalten die Mediator-Objekte

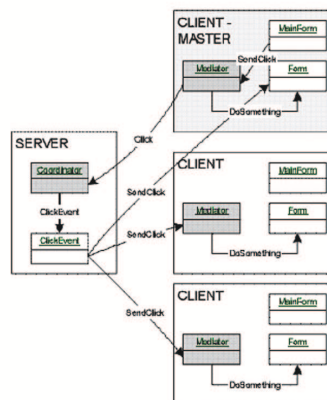


Abbildung 6.1: Schematische Darstellung des .NET Remoting Event Handling

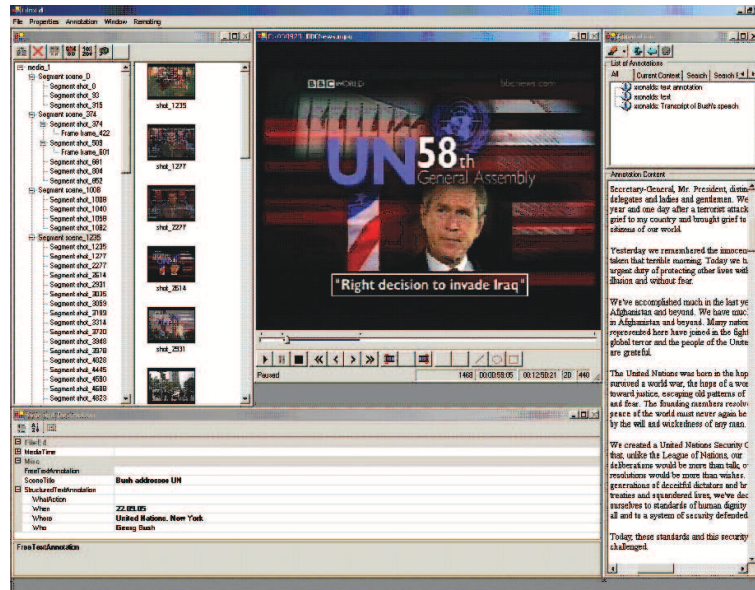


Abbildung 6.2: Benutzerschnittstelle des Vannotea Systems

die Kommunikation zwischen dem Client und dem Server. Der Koordinator kann so Methoden auf dem Mediator-Objekten aufrufen, welche Ereignisse auf allen anderen Clients Prozessen erzeugen. Hierbei ist das Ziel alle Ereignisse auf allen Clients zu simulieren. Auf diesem Prinzip werden auch alle Mausbewegungen simuliert. Die Maus jedes einzelnen Clients wird als „Pseudomaus“ auf den anderen Arbeitsstationen angezeigt. Hierbei werden die verschiedenen Pointer farblich eingefärbt um eine Unterscheidung zu ermöglichen.

Das „**MPEG2-Streaming**“ funktioniert auch auf Server Basis. Hierbei schickt der Server Befehle an den „Streaming“ Server, der die entsprechende Szene dann an die Clients streamt. Hierbei wird aus Effizienzgründen das IP Multicasting verwendet, um einer Datenflut entgegenzuwirken.

6.4 Benutzerschnittstelle des Systems

Die Benutzerschnittstelle (zu sehen in Abb. 6.2) besteht aus drei großen Einheiten:

- Content Player
- Content Description
- Annotation Discussion

Der „**Content Player**“ lässt Videoabspielmöglichkeiten (play,pause,suche,stop) zu und erlaubt auch die Annotation des gerade laufenden Frames, auf dem sich auch mittels Zeichenhilfen Regionen annotieren lassen. Der Inhalt der Annotation wird in der „**Annotation and Discussion**“-Komponente eingetragen. Hier

werden Details, wie Autor der Annotation und Zeitpunkt ebenfalls festgehalten. Die „**Annotation and Discussion**“-Komponente erlaubt dem Benutzer ebenfalls Annotationen zu durchsuchen und einzusehen. Im „**Content Description**“-Fenster können die einzelnen Szenen durchsucht werden.

6.5 Fazit

Dieses System ist mit dem Konzept einer kollaborativen, synchronen Echtzeit-Video-Annotation bisher einmalig und besitzt wegen seiner einzigartigen Architektur durchaus Potential in vielen Feldern eingesetzt zu werden. Bisher wird dieses System erfolgreich in Pilotprojekten in verschiedenen Bereichen wie dem „Australian Institute of Sports“, „CancerGrid“, „Indigenous Knowledge Management und dem „CyberSTEM - Networked Telemicroscopy“ eingesetzt. Gerade durch die Möglichkeit die Annotationsvorlage selbst den Anforderungen gerecht abändern zu können, ist dieses Tool in viele Bereiche erfolgreich vorgeedrungen. Der Meinung der Autoren zufolge werden noch weitere Erweiterungen des Systems folgen, wie die Möglichkeit, audiovisuelle Annotationen zu tätigen oder die Möglichkeit, durch eine interne Statistik das System und den Benutzer zu evaluieren. Auch der Schritt, statt audiovisueller Daten auch andere Datentypen zu unterstützen, könnte durchaus zu einer weiteren Verbreiterung des Einsatzspektrums dienen.

7 University of Indiana - eSports System

7.1 Ziele des eSports Systems(Gang Zhai u. Bulut 2004)

Das Hauptziel ist es ein kollaboratives synchrones Echtzeit Video-Annotationssystem zu schaffen, das über ein heterogenes Netzwerk funktioniert und eine Vielzahl an Benutzern zulässt. Hierbei soll der Einsatz im praktischen Sportannotations Bereich liegen, bei dem Trainer und Studenten bereits bestehendes Material oder Echtzeitübertragungen annotieren sollen. Ebenso soll das Systems durch Plugins erweiterbar sein.

Hervorzuheben ist, dass hier ein anderer Annotierungsansatz erfolgt als bei den bisher vorgestellten Systemen. Statt einer MPEG7-konformen Annotierung wird hier das Prinzip des „Media Content Enrichment“ verfolgt. Beim Content Enrichment werden grafische Zeichnungen, Text- und Audiodaten direkt in den ursprünglichen Video „Stream“ eingebunden und reichern somit das Eingangssignal mit Informationen an.

7.2 Übersicht über die Architektur des Systems

Dem System liegt eine „Message Brokering“-Middleware (zu sehen in Abb. 7.1) zugrunde (NaradaBrokering 2006), die verteilte Ereignisse zwischen einer großen Anzahl an Benutzern verteilen kann. Dabei ist die NaradaBrokering Middleware für sämtliche Kommunikation im System zuständig. Beim Narada-Brokering System handelt es sich um ein Multicast Broker MiddleWare-System auf dem Applikations Level, das viele Vorteile bietet, z.B. die Möglichkeit, Streams auch durch Proxys und Firewalls weiterzuleiten. Es unterstützt verschiedene Protokolle wie TCP, UDP, Multicast, SSL und RTP. Ebenfalls ist es möglich, das System in einem „Client-Server Mode“ zu betreiben oder in einem reinem „Peer-to-Peer Mode“.

Zusammen mit dem GlobalMMCS(GlobalMMCS 2006) System, das dafür zuständig ist, den Stream auf die heterogenen Endgeräte anzupassen, bilden diese beiden Technologien die Eckpfeiler des eSports Systems. GlobalMMCS basiert auf einem XGSP web-services Framework, hierbei integriert es Video-konferenzen, „Instant Messaging“ und „Real-time Video broadcasting“. Es setzt direkt auf die NaradaBrokering Middleware auf und bietet somit für jedes Endgerät das richtige Protokoll in der richtigen Bandbreite.

Dabei durchläuft das eingehende Live Signal zuerst das GlobalMMCS System und wird in Echtzeit auf einem „Storage“-Server abgespeichert. Von diesem wird das Signal an die Endanwender mittels eines RTSP-Servers weitergeleitet. Diese empfangen das Signal und können es mit Annotationen versehen, oder aus

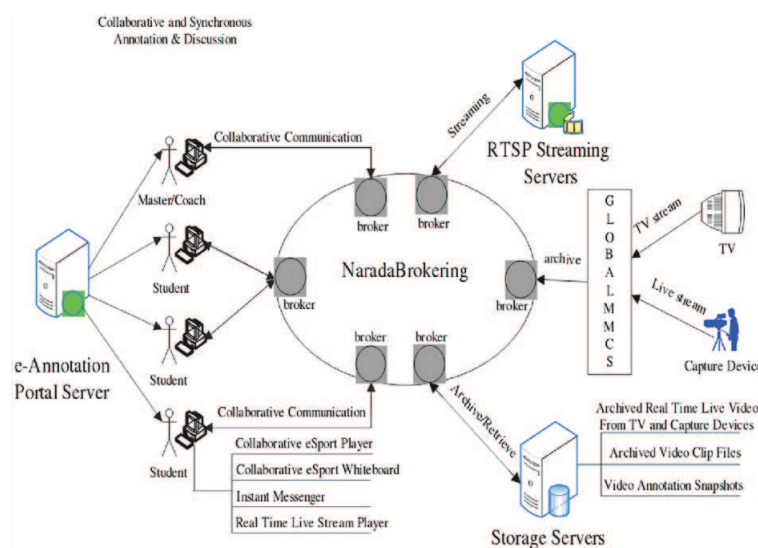


Abbildung 7.1: Schematische Darstellung der Architektur des eSports Systems

dem „Stream“ wichtige Screenshots generieren. Diese werden in Verbindung mit einem Whiteboard wieder an den Server übermittelt, der das Signal abspeichert. Von dort wird der „Stream“ mit den Annotationen angereichert, die wiederum in den „Playern“ angezeigt werden.

Die „Storage“-Server werden dezentral angelegt, wobei Daten auf jedem dieser Server abgespeichert werden können. Die darüber liegende NaradaBrokering Middleware sorgt dafür, dass der Benutzer keinen Unterschied sieht, wo die Information gespeichert wurde. Ein solches verteiltes System erhöht die Verfügbarkeit und Fehlertoleranz des Gesamtsystems.

7.3 Übersicht über die Methodik des Systems

- Alle Teilnehmer (Trainer und Studenten) können Echtzeit bzw. Archivaufnahmen anschauen und annotieren.
- Der Trainer kann Annotationen direkt auf dem Video „Stream“ tätigen und daraus einen „Verbundstream“ generieren, der mit dem original Videostream synchronisiert wird und diesen um die getätigten Annotationen ergänzt.
- Alle Benutzer können bereits annotierte und archivierte „Streams“ durchsuchen und abspielen.

7.4 Übersicht über die Komponenten des Systems

Das System besteht aus drei großen Komponenten:

- eSports Player

- eSports Whiteboard
- Instant Messenger

Der „**eSports Player**“ besteht aus vier Panelen: einem Panel, das die vorhandenen „Streams“ anzeigt, einem Panel, das Echtzeit-Video anzeigt, einem RTSP-Player und einem Video-Annotations-Whiteboard-Player. Das **Stream Panel** beinhaltet die Echtzeit-Videoliste, eine Liste an archivierten Videos und eine Liste mit bereits annotierten „Streams“. Das Echtzeit-Video-Panel zeigt aktuelle Live-Übertragungen, die vom Benutzer in der Liste ausgewählt wurden. Der **RTSP-Player** ist in der Lage Screenshots vom aktuellen Frame anzufertigen. Dieser Screenshot wird mit einem Zeitstempel versehen. Danach wird der Screenshot ins Whiteboard geladen, in dem es kollaborativ annotiert werden kann. Der eSports Player arbeitet kollaborativ, dh. jeder Benutzer sieht den gleichen Inhalt. Einzig der Trainer in seiner Rolle ist dazu in der Lage, das Video anzuhalten, zurückzuspulen und Screenshots anzufertigen.

Das „**Whiteboard**“ arbeitet genauso kollaborativ wie der eSports Player. Jeder Benutzer sieht den gleichen Inhalt, und Annotationen werden sofort bei allen anderen Benutzern sichtbar. Auf dem Whiteboard können sowohl Grafiken, als auch Text annotiert werden. Einzig der Trainer in seiner Rolle kann das Whiteboard abspeichern und es löschen.

Der „**Instant Messenger**“ wird benutzt, um das aktuelle Material zu besprechen. Jeder Benutzer kann Nachrichten schreiben und sieht alle Nachrichten, die von anderen Benutzern geschrieben wurden.

7.5 Fazit

Beim eSports System handelt es sich um einen Prototypen, der bisher zwar in vielen verschiedenen Universitäten auf der ganzen Welt angewendet wurde, wie z.B. bei der Kooperation der Beijing Sport University mit der Indiana University. Da es sich höchstwahrscheinlich um ein einmaliges Pilotprojekt handelt, wird das eSports System wahrscheinlich nicht ausgebaut werden. Als Prototyp-Projekt dient es als gutes Beispiel, wie synchrone Video-Annotationssysteme praktisch eingesetzt werden können. Hierbei ist besonders die gelungene Umsetzung der Echtzeit-Annotation hervorzuheben, die das System einzigartig macht. Sehr spezifisch ist auch die Lösung, bei der auf die NaradaBrokering Middleware und das GlobalMMCS-System aufgesetzt wird. Hier ist zu vermuten, dass diese Kombination der beiden durchaus als Basis für andere kollaborative Lösungen dienen kann und das eSports System nur eine dieser vielen Anwendungen ist. Negativ ist sicherlich festzuhalten, dass das System immer noch eine hohe Bandbreite braucht, um zu arbeiten, da es bisher nur mit dem MPEG1-Standard kompatibel ist. Sicherlich wünschenswert wären die Möglichkeit der Videokonferenz sowie einer Audio Annotation um eine bessere Kommunikation zu erreichen.

Literaturverzeichnis

ANNOTEA 2006

ANNOTEA: *Annotea Project*. <http://www.w3.org/2001/Annotea/>.
Version: 2006

Ching-Yung Lin u. Smith 2002

CHING-YUNG LIN, Belle L. T. ; SMITH, John R.: *Video Collaborative Annotation Forum: Establishing Ground-Truth Labels on Large Multimedia Datasets*. 2002

CLEF 2006

CLEF: *Cross-Language Evaluation Forum*. <http://www.clef-campaign.org>. Version: 2006

CWIRF 2006

CWIRF: *Chinese Web Test Collection*. <http://www.cwirf.org/>.
Version: 2006

DUBLINCORE 2006

DUBLINCORE: *The Dublin Core Metadata Initiative*. <http://dublincore.org/>. Version: 2006

Gang Zhai u. Bulut 2004

GANG ZHAI, Wenjun W. ; BULUT, Hasan: *eSports: Collaborative and Synchronous Video Annotation System in Grid Computing Environment*. University of Indiana, 2004

GlobalMMCS 2006

GLOBALMMCS: *Global Multimedia Collaboration System*. <http://www.globalmmcs.org/>. Version: 2006

IBM 2005

IBM: *Efficient Video Annotation (EVA) system*. <http://domino.watson.ibm.com/comm/research.nsf/pages/r.multimedia.innovation.html>.
Version: 2005

IBM 2006

IBM: *VideoAnnEx Annotation Tool*. <http://www.research.ibm.com/VideoAnnEx/>. Version: 2006

Microsoft 2006

MICROSOFT: *.NET Remoting Architecture*. <http://www.microsoft.com/germany/msdn/library/net/NETRemoting.aspx>. Version: 2006

NaradaBrokering 2006

NARADABROKERING: *The NaradaBrokering Project @ Indiana University*.
<http://www.naradabrokering.org/>. Version: 2006

NIST 2006

NIST: *National Institute of Standards and Technology*. <http://www.nist.gov/>. Version: 2006

Ronald Schroeter u. Kosovic 2004

RONALD SCHROETER, Jane H. ; KOSOVIC, Douglas: *Vannotea A Collaborative Video Indexing, Annotation and Discussion System For Broadband Networks*. University of Queensland, 2004

TimoVolkmer u. Apostol(Paul)Natsev 2005

TIMOVOLKMER, JohnR.Smith ; APOSTOL(PAUL)NATSEV: *A Web-based System for Collaborative Annotation of Large ImageandVideoCollections*. IBM T. J. Watson Research Center, 2005

TRECVID 2006

TRECVID: *TREC Video Retrieval Evaluation*. <http://www-nlpir.nist.gov/projects/trecvid/>. Version: 2006

Voorhees 2006

VOORHEES, Ellen: *Text Retrieval Conference*. <http://trec.nist.gov/>. Version: 2006

Eidesstattliche Erklärung

Ich versichere, dass ich meine Seminararbeit ohne Hilfe Dritter und ohne Benutzung anderer als der angegebenen Quellen und Hilfsmittel angefertigt und die den benutzten Quellen wörtlich oder inhaltlich entnommenen Stellen als solche kenntlich gemacht habe. Diese Arbeit hat in gleicher oder ähnlicher Form noch keiner Prüfungsbehörde vorgelegen.

Ort, Datum

Thomas Plotkowiak