

VIDEO ANNOTATION TOOLS

A Thesis

by

AHMED CHAUDHARY

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

May 2008

Major Subject: Computer Science

VIDEO ANNOTATION TOOLS

A Thesis

by

AHMED CHAUDHARY

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

Approved by:

Chair of Committee,	John J. Leggett
Committee Members,	Frank Shipman
	Steven E. Smith
Head of Department,	Valerie E. Taylor

May 2008

Major Subject: Computer Science

ABSTRACT

Video Annotation Tools. (May 2008)

Ahmed Chaudhary, B.S., Ghulam Ishaq Khan Institute of Engineering Sciences and
Technology

Chair of Advisory Committee: Dr. John Leggett

This research deals with annotations in scholarly work. Annotations have been studied by many people. A significant amount of research has shown that instead of implementing domain specific annotation applications a better approach is to develop general purpose annotation toolkits that can be used to create domain specific applications. A video annotation toolkit along with toolkits for searching, retrieving, analyzing and presenting videos can help achieve the broader goal of creating integrated work spaces for scholarly work in humanities research similar to existing environments in such fields as mathematics, engineering, statistics, software development and bioinformatics.

This research implements a video annotation toolkit and evaluates it by looking at its usefulness in creating applications for different areas. It was found that many areas of study in the arts and sciences can benefit from a video annotation application tailored to their specific needs and that an annotation toolkit can significantly reduce the time for developing such applications.

The toolkit was engineered through successive refinements of prototype applications developed for different application areas. The toolkit design was also guided by a set of features identified by the research community for an ideal general purpose annotation toolkit. This research contributes by combining these two different approaches to toolkit design and construction into a hybrid approach. This approach could be useful for similar or related efforts.

ACKNOWLEDGEMENTS

I would like to thank my committee chair, Dr. John Leggett, and my committee members, Dr. Frank Shipman and Dr. Steven Smith for their guidance and support throughout the course of this research. I would like to thank Dr. Du Li and Dr. Patrick Burkart for their helpful suggestions. And I would like to thank my mother and father for their encouragement and love.

TABLE OF CONTENTS

	Page
ABSTRACT	iii
ACKNOWLEDGEMENTS	v
TABLE OF CONTENTS	vi
LIST OF FIGURES	viii
1. INTRODUCTION	1
1.1. Problem Statement	2
1.2. Hypothesis	3
1.3. Proposed Solution	3
1.4. Methodology	4
1.5. Method of Evaluation	4
1.6. Expected Results	5
2. LITERATURE REVIEW	7
2.1. Video Annotation	7
2.2. Video Annotation Application Scenarios	10
2.3. Related Work	11
2.4. Comparison of Recent Tools	21
3. ANATOMY OF THE TOOLKIT	22
3.1. Core Interfaces	22
3.2. Implementation Classes	24
3.3. User Interface Controls	25
4. EVALUATION	27
4.1. Film Studies	27
4.2. Medical Sciences	31

	Page
4.3. Scholarly Texts	32
4.4. Summary.....	35
5. CONCLUSIONS	36
5.1. Comparison with an Ideal Toolkit	36
5.2. Future Directions.....	39
REFERENCES	41
VITA	45

LIST OF FIGURES

	Page
Fig. 1. Marquee User Interface	13
Fig. 2. The XMAS User Interface	15
Fig. 3. XMAS Discussion Forum	16
Fig. 4. The DIVER User Interface	17
Fig. 5. The LEAN System Running on a TabletPC	18
Fig. 6. Prototype in Authoring Mode	29
Fig. 7. Prototype in Playback Mode	30
Fig. 8. Annotated Sonogram Video	32
Fig. 9. Annotated Document	34

1. INTRODUCTION

A toolkit for developing collaborative video annotation applications for humanities research is proposed. A video annotation application provides a means of marking up the objects and scenes in video streams in order to facilitate interpretation and understating of content [11].

Broadly speaking multimedia can be annotated in two different ways, by metadata association and by content enrichment as described by Gang et al. [11]. Metadata association uses specific metadata models to build a semantic structure which supports operations such as search. This approach requires the user to understand the underlying semantic metadata model in order to perform annotations that conform to the framework. Moreover, the user must spend much time and effort in marking the multimedia object, which is a tedious task [11].

The second method of content enrichment uses other multimedia elements such as graphic shapes, text, and audio to enrich the multimedia objects in a multimedia stream and generate a new composite stream. Users interactively add annotations by associating text, lines, rectangles and other shapes to frames or clips of a video. In this way, the original stream's content is enriched by the new content. This method is more straightforward for viewers to understand and can be used in a collaborative way [11].

This thesis follows the style of *IEEE Transactions on Software Engineering*.

1.1. PROBLEM STATEMENT

A simple generic model of humanities research is presented by Marsden et al. [24]. This model views humanities research as a process of repeatedly accessing, searching, annotating, transcribing, analyzing, and presenting materials. The order of these operations varies. Scholars constantly cycle between different ways of working with audiovisual materials. Innovative research often combines them in unexpected variations or applies them to different materials [24]. Researchers often need to collaborate with each other at each step. Hence there is a need for tools that integrate these activities, provide facilities for collaboration and provide flexible mechanisms for performing these activities in various combinations. These toolkits can then be used to rapidly create applications and workflows to facilitate different needs of humanities researchers.

This report focuses on the annotation of video and proposes a research endeavor to develop a general purpose collaborative video annotation software toolkit. Such a toolkit would not only allow creation of applications that allow individuals to annotate materials but would also make it possible to create applications that allow easy sharing of annotations. This would allow researchers to collaborate over different types of media. Annotation of audiovisual materials can take a lot of time, and even if material has been annotated by one researcher, the problem remains of how other researchers can make use of the annotation. Annotation for the purpose of finding audiovisual material seems successful, but we have not seen anything like the sophisticated and consistent analysis that would be needed to write even a basic film or book review [24].

1.2. HYPOTHESIS

Grudin and his colleagues have worked on multimedia annotation systems for a number of years and have created some prototype systems for different domains [2, 3, 4, 5, 16, 20, 21]. However after three years of research and system building in this area they have concluded that rather than building a general purpose multimedia application it is more beneficial to build a toolkit and relevant infrastructure that allows developers to create video annotation tools specific to users' needs. Such a toolkit along with toolkits for other research activities like accessing, searching, presenting and analysing video material will move us closer to the development of integrated analysis environments for humanities research [24].

1.3. PROPOSED SOLUTION

It seems that for the time being there is no single tool that would satisfy a wide variety of different scenarios where multimedia must be annotated collaboratively. Some tools focus on synchronous collaboration, some focus on asynchronous collaboration and yet others focus on interesting interaction techniques for creating annotations. This trend leads the author to believe that it is not a good idea to implement a general purpose tool that would attempt to satisfy all multimedia annotation needs. Instead it would be better to develop a toolkit that can be used for creating annotation applications for different domains.

1.4. METHODOLOGY

One method of developing a software toolkit is to first implement a sample application(s) that provides functionality similar to the toolkit being developed and then generalize reusable parts of the application(s) into a toolkit. One such methodology for developing toolkits, based on iterative refinement of light weight prototypes, is proposed by Edwards et al. [10]. The author recommends this methodology to develop the proposed toolkit.

In order to implement a generalized toolkit for collaborative multimedia annotation we need a conceptual framework to represent annotations and objects that will be annotated. One such framework has been presented by Barger and his colleagues [1]. The framework allows any type of media to be annotated with any other type of media and has been designed to be flexible, extensible and platform independent.

1.5. METHOD OF EVALUATION

Evaluating a software toolkit is different from evaluating a piece of software because a toolkit is required to support many different scenarios where as a single piece of software is required to satisfy requirements for a specific scenario and set of users. Edwards, et al. [10] describe some methods and guidelines for a user-centered evaluation of toolkits.

According to Edwards, et al. [10] evaluation via lightweight technology prototypes is a good method to evaluate the quality of software toolkits. This approach involves the creation of a number of lightweight “throw away” applications, purely to

demonstrate the utility of novel aspects of the toolkit. These lightweight “throw away” applications, while engineered to be neither robust nor feature-full enough for long-term use, are easy to build and require few engineering resources [10].

The lightweight prototypes can be evaluated by humanities researchers and they can give their feedback to the developers who can in turn improve the toolkit and create more light weight applications that demonstrate newer functionality in the toolkit. The feedback can be gathered using usability studies and experiments, questionnaires and interviews.

In this way the toolkit can be developed iteratively and evaluated after each iteration until it provides most of the functionality deemed important by the researchers and the developers. The developers’ feedback would also be important because ultimately the toolkit would be a set of software components and APIs (application programming interfaces) that would be used by other developers to create domain specific video annotation applications for the researchers. The developers would be interested in the ease of use and modularity of the APIs and components.

1.6. EXPECTED RESULTS

The author expected the research to produce a toolkit that allows developers to rapidly develop video annotation applications suited to different needs. This in turn will allow researchers to share their views with others. Grudin and Barger [16] et al. distilled the following requirements for a generic multimedia annotation platform to

support diverse asynchronous collaboration scenarios. It is hoped the developed toolkit satisfies these requirements:

1. Thorough support for common activities:

Common annotation functions like: creating, saving, retrieving, and deleting annotations should be the easiest to incorporate into an interface.

2. Extensibility and customizability:

The toolkit should provide extensibility and customizability at both the user interface and functionality levels.

3. Storage flexibility:

Designers should be able to store annotations in a variety of configurations.

4. Universal annotation support:

A general-purpose annotation toolkit should support annotating any media type with any other media type.

5. Interoperability:

Task specific user interfaces should be interoperable i.e. annotations made in one user interface based on the toolkit should be transferable to another user interface based on the toolkit with minimal effort.

2. LITERATURE REVIEW

2.1. VIDEO ANNOTATION

The metadata associated with a resource can be sufficient for locating a resource, but once a resource is found; there is often the need to associate finer grained metadata with certain points within the audiovisual content. This potentially rich process is what we call annotation.

It is often very time consuming to make annotations, so the challenge is to allow users to do so in a way that has some enduring value. One response is the development of standards to render annotations durable and facilitate their reuse by others. Important developments in this area are MPEG-7 and Annodex, but neither has as yet been widely adopted. Collaborative annotation systems provide another approach to durability of annotations, by establishing a form of consensus and saving effort by involving more users [24].

In the context of time-based media, annotation associates extra information, often textual but not necessarily so, with a particular point in an audiovisual document or media file. In humanities research, annotation has long been important, but in the context of sound and image, it takes on greater importance. Rich annotation of content is required to access and analyze audiovisual materials, especially given the growing quantities of this material. Annotation software for images, video, music and speech is widely available, but it does not always meet the needs of scholars, who annotate for

many different reasons. Sometimes annotation simply allows quick access or index of different sections or scenes. Annotation has particular importance for film and video where annotation is sometimes used for thematic or formal analysis of visual forms or narratives. At more fine grained levels, some film scholars analyze a small number of film frames in detail, following camera movements, lighting, figures, and framing of scenes. Annotation tools designed for analysis of cinema are not widely available. Most video analysis software concentrates on a higher level of analysis [24].

2.1.1. Video Annotation and Standards

Many different approaches exist for standards in annotation. Several well-known metadata standards are applicable to humanities research, such as MARC [22], and Dublin Core [9]. These are useful standards, but are dominated by the resource level approach; most similar metadata standards describe content on the level of an entire entity within a library. This level of metadata is very useful, but does not satisfy the requirements of annotation for video. These standards do not have robust models for marking points *within* the content [24].

MPEG-7 is an ISO standard conceived in 1996, and finalized in 2001-2002. It is intended to be a comprehensive multimedia content description framework, enabling detailed metadata description aimed at multiple levels within the content.

Technically, MPEG-7 offers a description representation framework expressible in XML. Data validation is offered by the computationally rich, but somewhat complex XML Schema standard. Users and application providers may customize the precise

schema via a variety of methods. Numerous descriptive elements are available throughout the standard, which can be mixed and matched as appropriate. Most significantly, it allows for both simple and complex time and space-based annotations, and it enables both automated and manual annotations [24].

Another standard for annotation of video content that draws on MPEG-7 is the Annodex [28]. It is an open standard for annotating and indexing networked media. Annodex tries to do for video what URL/URI (i.e. web links) have done for text and images on the web. That is, to provide pointers or links into time-based video resources on the web. The Metavid project [25] demonstrates Annodex in action on videos of the U.S. Congress [24].

Both Annodex and MPEG-7, rich as they are, do not support other types of media such as text and 3d models as targets of annotations.

2.1.2. Collaborative Annotation

A number of projects have attempted to design and construct collaborative software environments for video annotation. In collaborative video annotation, a number of people can work on the same video footage. Efficient Video Annotation [34] is a novel Web tool designed to support distributed collaborative indexing of semantic concepts in large image and video collections. Some video annotation tools such as Transana [38] have multi-user versions in addition to single-user versions [24].

2.2. VIDEO ANNOTATION APPLICATION SCENARIOS

Schroeter et al. [29] give some examples of areas in the sciences where video annotation is useful e.g. oceanographic studies, crystallography and biology. Some scenarios where the proposed toolkit would be useful are described below.

2.2.1. Film Studies

In film studies teachers often have to point out subtleties in a movie to students. If teachers were able to annotate a video with their comments, then the students could more easily understand these subtleties.

Students of film studies are often required to create video-based works. Traditionally the instructor evaluates the work and gives feedback to the students either verbally or in written form separate from the work itself. A video annotation system would allow the instructor to give feedback on the work itself. Various authors [7, 12, 18] give us an insight into what types of tasks are usually performed by film scholars in their studies.

2.2.2. Biology

Many subfields of biology produce video as research output. From video of wild animals in their natural habitat to videos of observations made through microscopes, biologists frequently need to study videos collaboratively to exchange ideas and transfer knowledge. Annotation of videos can help them perform these tasks efficiently.

2.2.3. Coaching and Teaching

Teaching and coaching in various fields, like surgery and performing arts, requires trainers to comment on students' performances. Video is frequently used as a means of recording the performance and commenting on it. For example coaching of football players requires that they review video recordings of their earlier performances. These videos are discussed and critiqued by the coach. One such system is described by Gang et al. [11]

Another example is learning dance composition which requires students to study their own performance and learn from their mistakes. It also requires teachers to give feedback on students' performances. Video annotation can make these tasks much easier and intuitive for both the teachers and dancers. Gina et al. [6] describe a system which uses video annotation to help dance composition students.

Goh et al. [13, 14] report a study conducted with political science students where they were required to study political speeches and author a presentation on the subject matter. The study used a tool called Synchrony. Synchrony was created as part of research into patron augmented digital libraries. In addition to other digital libraries related functions it allows the users to associate text elements with video to author mixed text and video presentations.

2.3. RELATED WORK

A number of systems [8, 11, 15, 27, 29, 30] have been developed over the years that allow manual annotation of video. Some of these systems also allow collaboration

scenarios around annotations. The type of collaboration facilitated by these tools varies. Some tools do not have any collaboration features at all—annotations are meant for private use only, some allow asynchronous collaboration, and others allow synchronous collaboration. A brief overview of some of these systems is provided below.

2.3.1. Early Tools

In the late 80s and early 90s a number of tools were developed, that provided basic annotation facilities to researchers working with video. Harrison et al. [17] identified the requirements for such tools and developed a prototype application called VANNA [17]. Other tools from this time period are described in the following subsections.

2.3.1.1. Videonoter

The Videonoter [33] application used hypermedia links to relate video segments to annotation text or graphical annotations in an editor. It also allowed the users to focus on specific areas of interest in the video. The user interface provided time-ordered columns that allowed side-by-side comparison of data.

2.3.1.2. EVA

The Experimental Video Annotator (EVA) tool was created by Mackay [23] to help researchers analyze video for interesting events. EVA allowed users to associate tags and text-based comments with video segments and allowed them to capture and

annotate specific frames of the video. It allowed the users to navigate through the video using the associated tags.

2.3.1.3. Marquee

Marquee was a tool created by Weber et al. [35] in the mid 90s. Marquee was intended to assist people in accessing information recorded on a videotape.

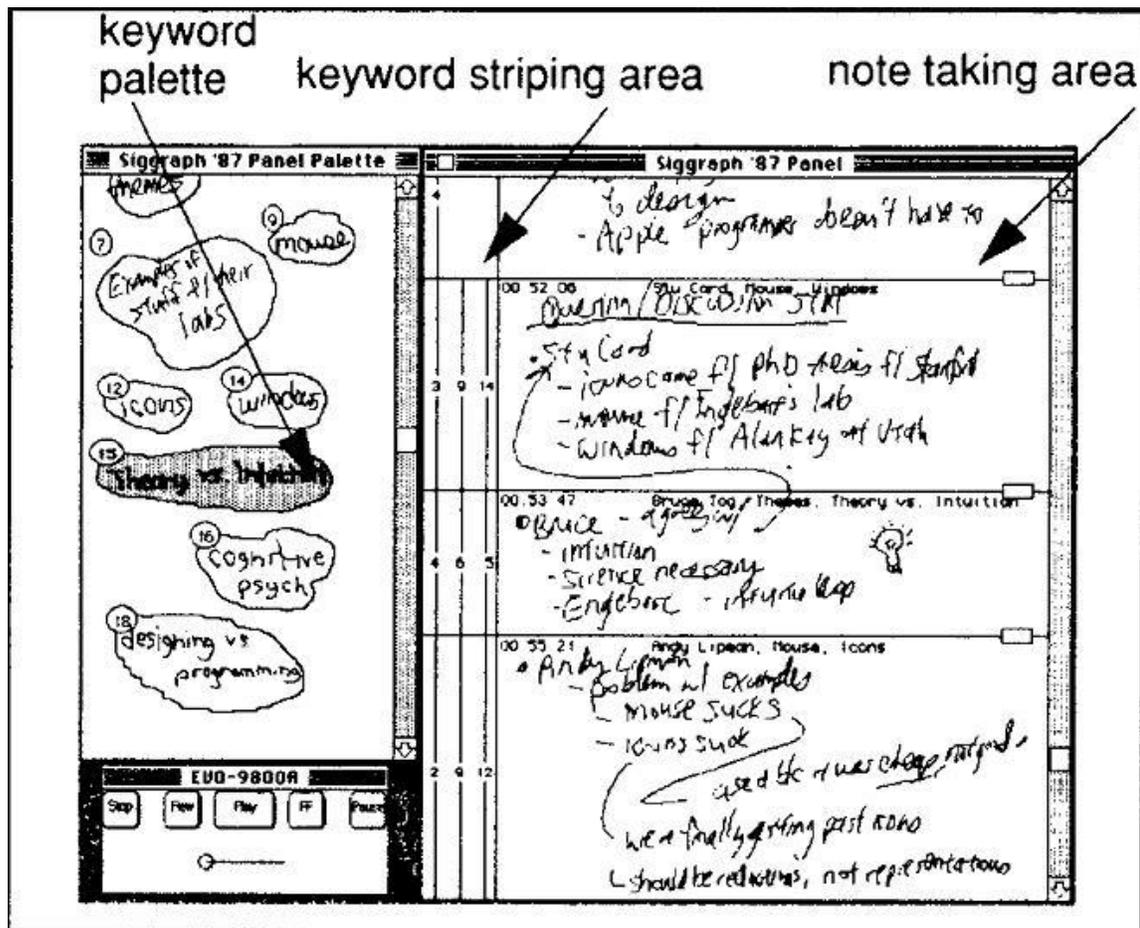


Fig. 1. Marquee User Interface

It was a pen-based system which provided users with a means for correlating their personal notes and keywords with recorded footage [35]. Figure 1 shows the Marquee user interface.

2.3.2. Recent Tools

Some of the more recent video annotation tools are described below in the following subsections.

2.3.2.1. MediaMatrix

Media matrix [19] is a web-based tool that allows users to collect material from the web and add annotations to it. It also allows the users to create presentations with the materials they have gathered and annotated. MediaMatrix uses browser plug-ins to provide its functionality. It is a completely server-based application and does not store anything on the client side. It allows users to select different media types like images, text, audio and video from a web page and add them to the user's collection. The system only allows text-based annotations on the different media types.

2.3.2.2. XMAS

XMAS— the Cross Media Annotation System — provides tools to enhance the use of video and image collections in humanities courses and in any subject in which precise reference to visual materials is needed. XMAS can be used in conjunction with

image and text collections, and is currently optimized for use with commercially available DVDs as video source. Figure 2 shows the XMAS user interface.

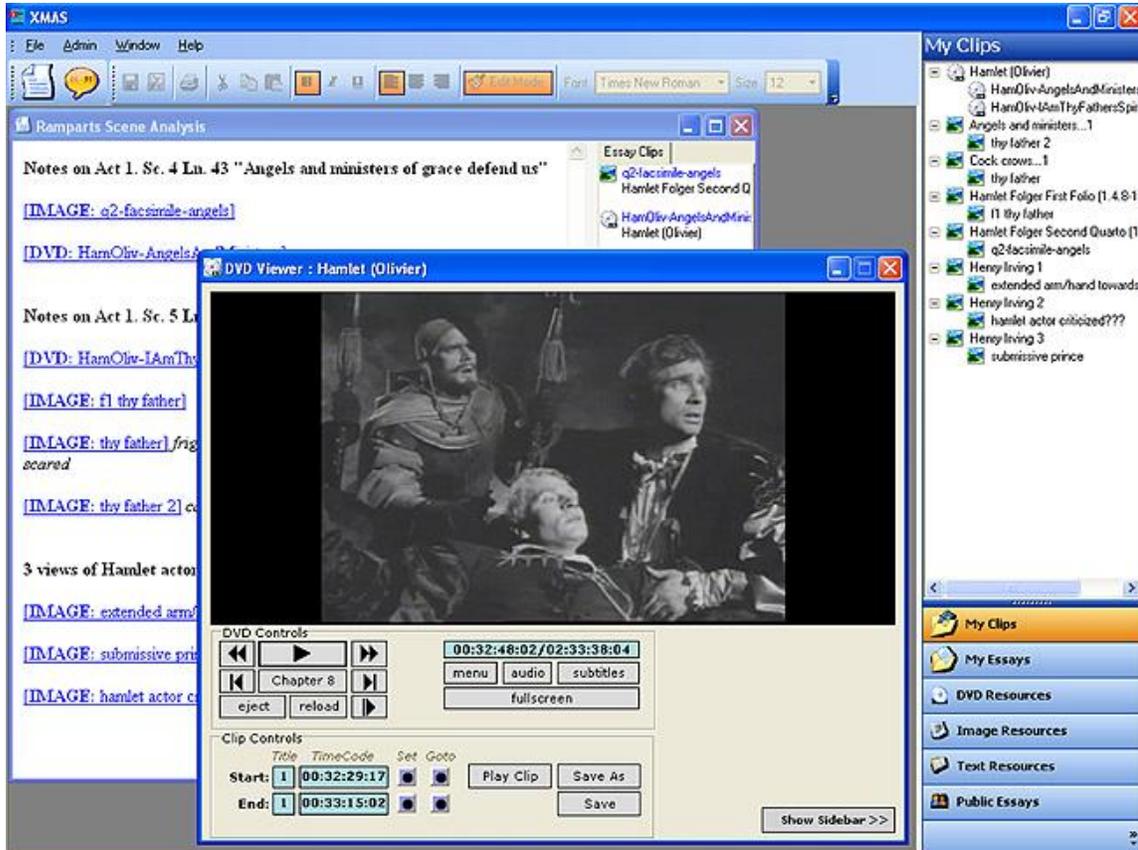


Fig. 2. The XMAS User Interface

XMAS allows users to rapidly define segments of film which can be replayed by clicking on automatically created links that can be saved in a list or dragged and dropped into discussion threads or online essays. It allows students to select, annotate and share video sequences for use in on-line discussions, multimedia essays and in-class presentations [8]. Figure 3 shows the XMAS discussion forum user interface.

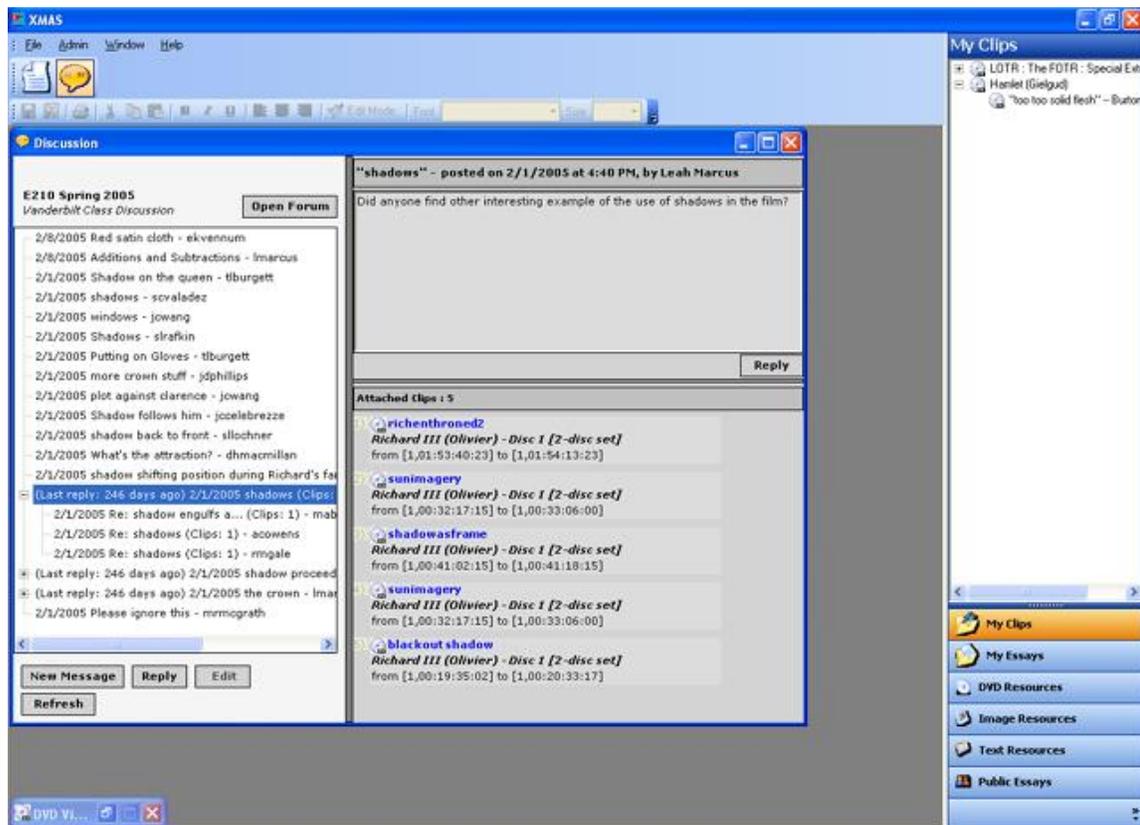


Fig. 3. XMAS Discussion Forum

2.3.2.3. DIVER

DIVER (digital interactive video exploration and reflection) [27] is a tool for authoring and sharing DIVES. A DIVE is an annotated perspective on any video record. Content can be captured by equipment ranging from basic consumer video cameras to specially built, high-resolution 360-degree panoramic cameras with a multi-microphone array. DIVER allows for infinite points-of-view and commentary from a single video

recording. Desktop DIVER allows users to import source movies and create new annotated "paths" through the video source. The new annotated movie is the user's own personal DIVE. WebDIVER allows DIVERs to upload a DIVE and share it with others who, in turn, can comment on the DIVE. Figure 4 shows the DIVER user interface. The overview window (bottom left) shows the full video source. The magnified viewing window (upper left) shows a selected image from the scene. The annotation window, or the Dive worksheet (right) lets users comment on the frames or path movies they create.

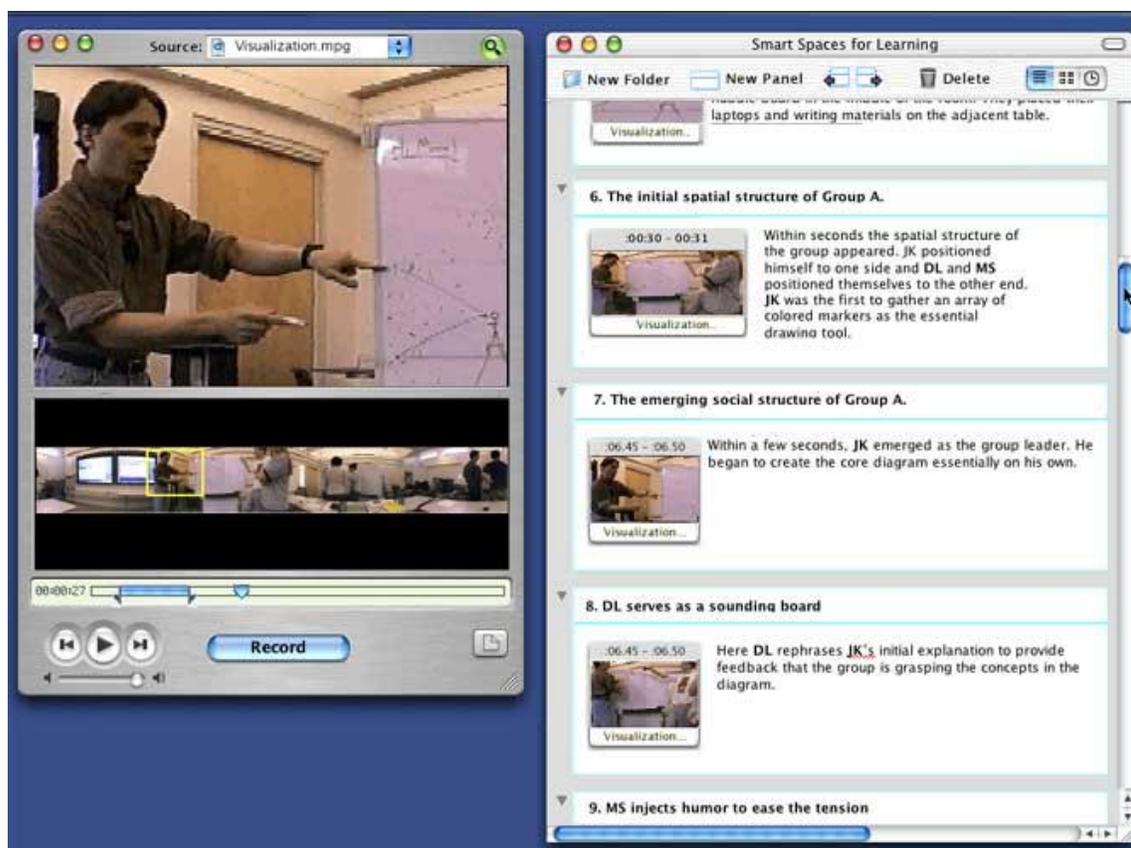


Fig. 4. The DIVER User Interface

2.3.2.4. LEAN

Ramos and Balakrishnan developed a system [15] called LEAN that serves as an exploratory platform for new visualization and fluid interaction techniques for navigating and controlling digital video. Their system targets the casual user, and in addition to various editing operations, allows for casual annotation and cross-linking of video streams. Its primary interface is a digitizer tablet with a pressure-sensitive pen. Their intention is to leverage users' familiarity with pen-based interactions in the physical world, and emerging tablet-based computers [15]. Figure 5 shows the LEAN system running on a TabletPC.

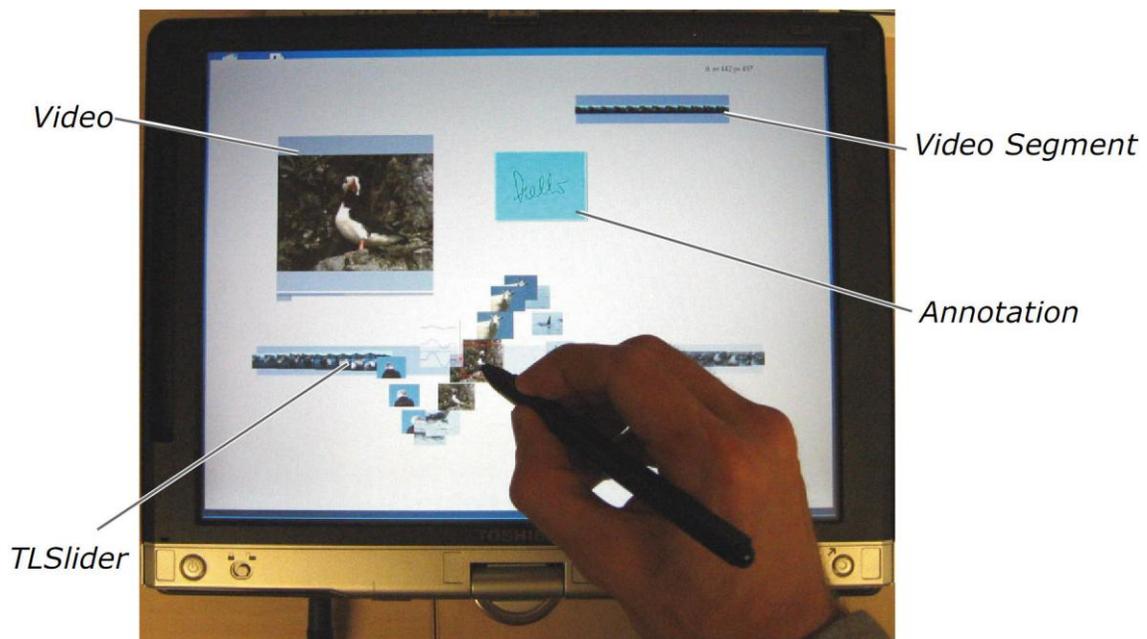


Fig. 5. The LEAN System Running on a TabletPC

The LEAN system allows for the manipulation of a video stream by using a small set of gestures that lets users start, stop, and travel to any arbitrary point in time in the stream. Also, by using only simple gestures, users are able to select intervals, or segments, from the video. The system provides a novel interaction and visualization technique based on fish-eye lenses called the Twist Lens slider or the TLslider. The TSslider provides a visualization of the complete video stream as a sequence of thumbnails.

Besides allowing users to manipulate the video stream, the system also permits users to attach annotations – easily created by scribbling on the working area or over the video image – to video frames and segments. By connecting an annotation to a desired element on the working area, the user can provide it with a positional and temporal context. In addition, users can trigger visualizations that correspond to a complete video segment and that also allow for both the quick navigation of the video stream and the speedy location of the annotations situated within [15].

2.3.2.5. Vannotea

Vannotea [30, 31] is a prototype system developed at the Distributed Systems Technology Centre, at the University of Queensland, as part of the FilmEd project, which enables the real-time collaborative indexing, browsing, description, annotation and discussion of high quality digital film or video content. It supports large-scale group-to-group collaboration. Users can open an MPEG-2 file and share the tools which enable the group to collaboratively segment, browse, describe, annotate and discuss the

particular film or video of interest. The descriptions and annotations can be shared by saving them to a server [31].

2.3.2.6. Transana

Another video annotation tool is Transana [38] developed by University of Wisconsin, which allows researchers to ‘identify analytically interesting clips, assign keywords to clips, arrange and rearrange clips, create complex collections of interrelated clips, explore relationships between applied keywords, and share their analysis with colleagues.

2.3.2.7. VAST

Video Annotation and summarization tool (VAST) was developed by Mu and Marchionini [26]. VAST integrates both semantic and visual metadata. The system generates candidate key frames by selecting every n th frame from the video. The user selects interesting frames (key frames) from these and then associates textual metadata or annotations with them. The system can also generate fast forward surrogates for the user and the user can associate textual metadata or annotations with these surrogates.

The system combines the visual metadata (key frames and fast forward surrogates) and the semantic metadata (textual metadata associated with the key frames and surrogates by the user) into an XML file. This XML-based data can be used by other applications. VAST was originally developed to create fast forward surrogates for the Open Video Library.

2.4. COMPARISON OF RECENT TOOLS

These tools provide different degrees of support for collaboration and annotations on video. The XMAS [8] system has strong features for the creation of multimedia documents based on DVD versions of films. But it is limited to DVDs only as the source of video. The DIVER [27] tool provides a unique way of providing different perspectives on the same video. However it does not support annotating video with media other than text. MediaMatrix [19] and VAST[26] have this same limitation. The Vannotea [30, 31] system provides more collaboration features than any other tool discussed. It provides for synchronous collaborative annotation features in addition to the basic features provided by the other systems. However, it does not allow annotating a video with another video. The LEAN [15] system provides some very interesting interaction techniques for browsing and annotating video but does not seem to have any collaborative features.

3. ANATOMY OF THE TOOLKIT

This section describes the toolkit construction. As mentioned earlier it was decided to construct the toolkit using successive refinements of prototype applications. These scenarios are described in the next section. This section describes the various classes, interfaces and user interface elements that make up the toolkit.

It was decided that the system will use standards-based technologies like XML to save and share the annotations and that it would be built using C# and Microsoft .NET Framework 3.0 [36] technologies. The choice of platform was influenced by the author's previous experience with the .Net framework.

This toolkit was implemented using a simple iterative object-oriented software development process. The tools used were Visual studio 2005 and Windows Presentation Foundation [37]. The toolkit code can be divided in three logical pieces core interfaces, classes that implement these interfaces and user interface elements. Some of the important classes, interfaces and user interface controls are described below in the following subsections.

3.1. CORE INTERFACES

The toolkit was designed to support extensibility. This is accomplished by providing hooks into the toolkit so that developers can provide custom implementations for parts of the toolkit if needed. The toolkit itself implements these interfaces to provide

its functionality. These hooks take the form of C# interfaces. Some of these are described below in the following subsections.

3.1.1. IAnnotationService

This interface defines the functionality required to show annotations on an instance of `IVideoAnnotationsHost`. It uses an instance of the `IVideoPlayer` to interact with the target video and it uses an instance of the `IAnnotationStore` to save and load annotations.

3.1.2. IVideoAnnotationsHost

Instances of this interface are responsible for interacting with the user for creating annotations and displaying the target video and annotations on the screen. This interface uses an `IVideoPlayer` instance to interact with the target video. By providing custom implementations of this interface different annotation layout schemes can be implemented. The default implementation provided by the `MediaWorkspace` class in the toolkit uses a desktop like layout for the annotations.

3.1.3. IVideoPlayer

Instances of this interface are responsible for playing a single video on the screen. This interface allows the developer to use a custom video player, if the toolkit provided video player does not meet the needs.

3.1.4. IAnnotationStore

Instances of this interface are responsible for saving and loading annotations from a persistence mechanism. The toolkit provides a default implementation of this interface `XmlStore` which saves and loads annotations from the file system. This interface allows the developer to provide custom storage mechanisms for the annotations e.g. this interface can be implemented such that the annotations are stored in a relational database instead of the default XML and file system based mechanism provided by the toolkit.

3.2. IMPLEMENTATION CLASSES

Some of these classes implement the core interfaces described above to provide the toolkit functionality and others provide support to these classes.

3.2.1. Annotation

This is the base class for all different annotation types. Classes representing different annotation types derive from this class. Instances of this class represent a single annotation. Instances contain the annotation contents. Currently it supports two types of contents; plain text and rich formatted text. This class keeps track of location within a target where the annotation was placed. Instances of this class can provide an XML representation of themselves for persistence purposes. Some of the classes that derive from this class include `InkAnnotation`, `VideoAnnotation` and `RichTextAnnotation`.

3.2.2. XmlStore

This class implements the IAnnotationStore interface and is responsible for saving and loading annotation from the file system.

3.2.3. CanvasAnnotationService

This class implements the IAnnotationStore interface. It is responsible for initializing the annotation system. It keeps track of all the annotations made on a video. It interacts with the user interface layer classes to display the annotations on the user interface and to get the annotation content from the user interface. It also saves and loads annotations from XML files using the XmlStore class.

3.2.4. AnnotationCollection

This class represents a collection of annotations and is used by the XmlStore class to save and load annotations.

3.3. USER INTERFACE CONTROLS

These classes represent user interface controls.

3.3.1. SimpleVideoPlayer

This class implements the IVideoPlayer interface. This is a user interface layer class that is responsible for playing an individual media file. It provides control over the media playback.

3.3.2. MediaWorkSpace

This is a user interface level class that is responsible for hosting instances of the `IVideoPlayer`. It allows the users to resize and drag the videos on the screen. The annotation service class interacts with this class to display and hide the annotations during media playback.

3.3.3. WebCamCaptureControl

This control is responsible for recording video from a webcam.

3.3.4. InkAnnotationControl

This control is responsible for showing ink annotations on the screen.

3.3.5. RichTextDisplayControl

This control displays text annotations and allows formatting of the text.

3.3.6. Duration Control

This control allows the user to set the annotation duration visually by scrubbing through the target video.

4. EVALUATION

The toolkit was evaluated and built by prototyping scenarios where applications based on the toolkit would be useful. The first scenario considered was film studies. The second scenario was annotating medical videos for teaching and consultation purposes. And the third scenario was annotating scholarly documents for critiquing and enriching the content of documents. These prototypes and scenarios are described below in the following subsections.

4.1. FILM STUDIES

Students of film are required to write film analysis essays as part of their studies. Currently these are written in word processors with relevant frames embedded in the text. The problem with this approach is that often the analysis deals with movement of the camera, changes in lighting, actor performances in a certain scene etc. and highlighting these things with static images is difficult. There is a need to be able to associate critical pieces of writing with the video itself. To explore these scenarios a prototype was implemented.

The aim was to implement a sample application that allows film scholars to annotate videos in a collaborative fashion. Based on a quick study [7, 12, 18] of how film students perform their scholarly work the author came up with a set of requirements for the prototype. It should allow the user to:

- Annotate frames, shots etc. The annotations can be text based annotations, free form hand drawn annotations that can be put on top of the video or video clips captured with a webcam.
- Share annotations with others.
- Compare one film or clip with another.

The prototype that was implemented allows the user to create plain text, rich text, video clip and digital ink based annotations which can be associated with parts of a video. It also allows the user to view multiple videos side by side. The user interface allows the user to freely move the playing videos and annotations on the screen and to resize them. The user can zoom in and zoom out of the entire workspace containing the video and annotations as well. The prototype can save the annotations as a zip file that contains an xml file along with any resources that were created as part of the annotation process (e.g. a web cam captured video clip).

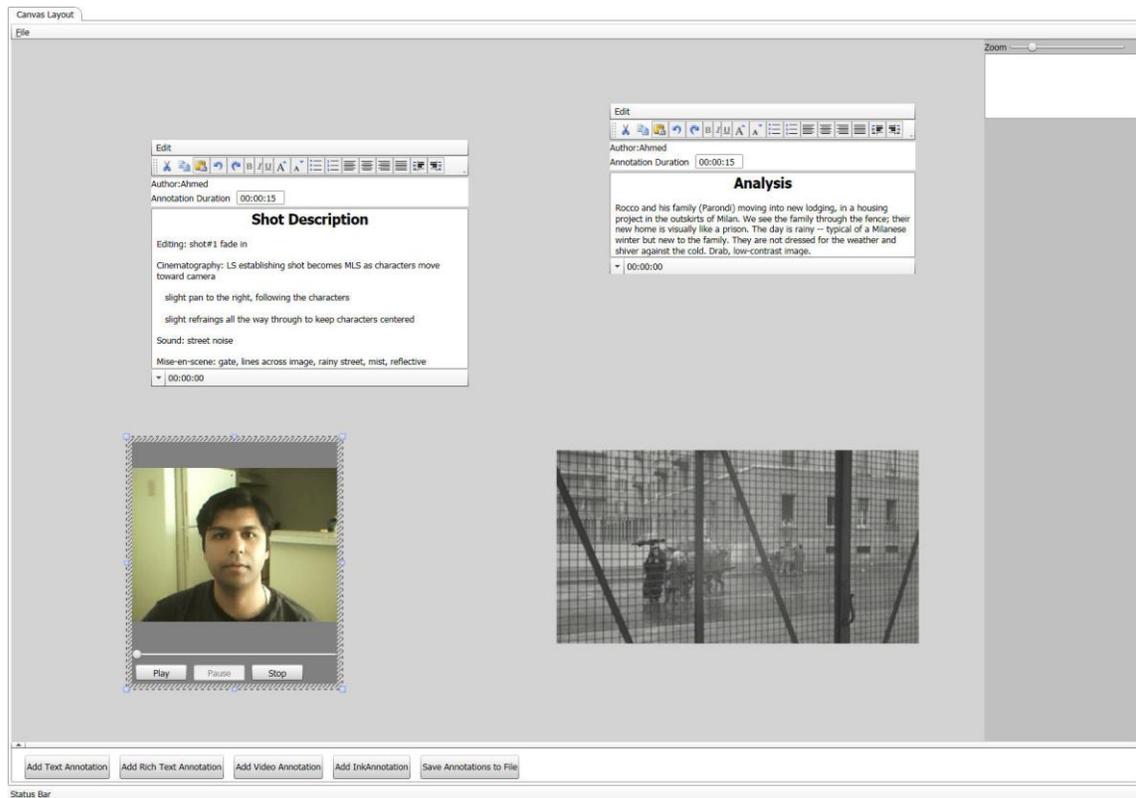


Fig. 6. Prototype in Authoring Mode

This zip file can be shared with other users and these users can add their annotations to the same video. In this way the prototype allows users to collaboratively annotate a video in an asynchronous fashion. Figure 6 shows the user interface of this prototype in authoring mode.

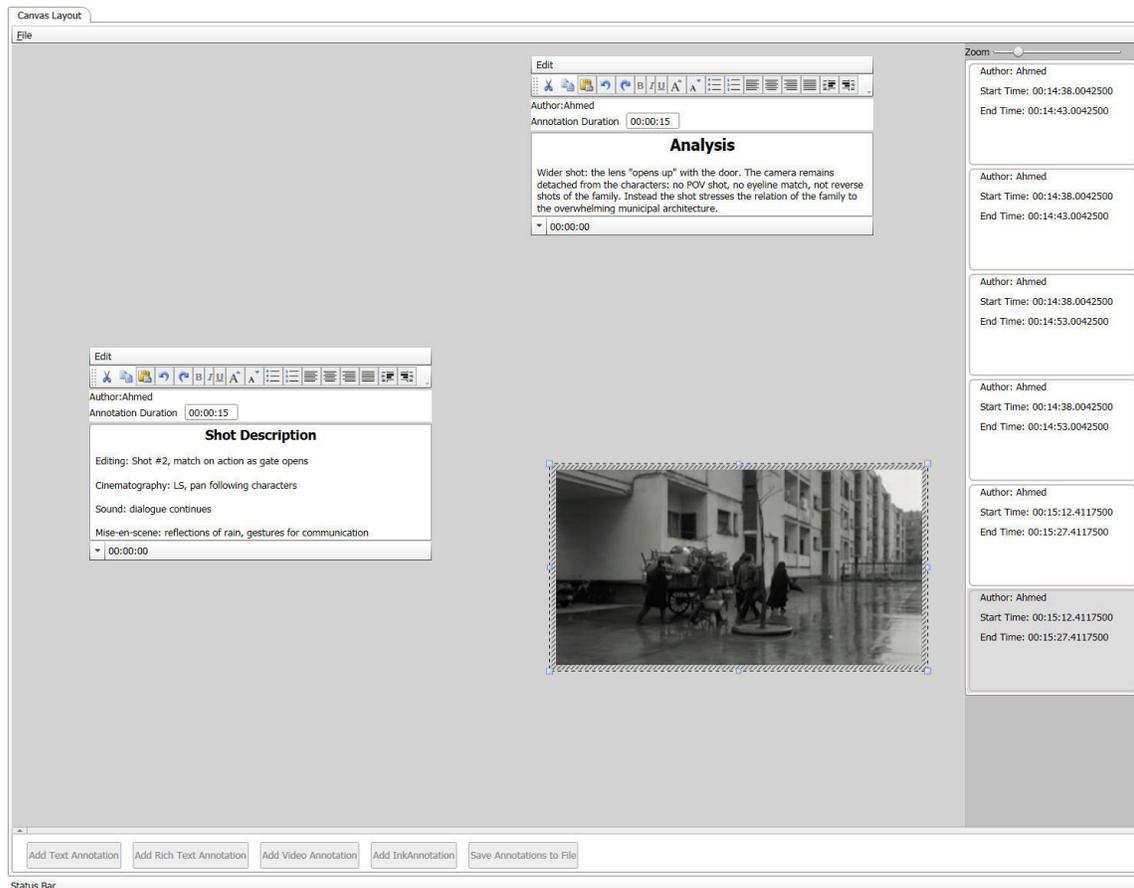


Fig. 7. Prototype in Playback Mode

The figure shows two text based annotations and a web cam captured video annotation along with the target black and white video. The annotations and the target video make up the workspace area. The buttons at the bottom of the figure allow the user to add different types of annotations. The slider control in the upper right corner allows the user to zoom in and out of the workspace area.

Figure 7 shows the prototype in playback mode where the user is viewing an annotated video. The user interface is very similar to the authoring mode except for the list view panel on the right hand side which shows the annotations and allows jumping to the point in the video where the annotation was added.

4.2. MEDICAL SCIENCES

Doctors sometimes have to analyze videos as part of their work either for diagnosis or for teaching and consulting purposes. To explore this area the film studies prototype was enhanced to support ink on video annotations. Figure 8 shows how a video can be annotated by a doctor for teaching purposes. It shows a sonogram video of a fetal leg with a digital ink annotation on top of it and text and video annotations by its sides.

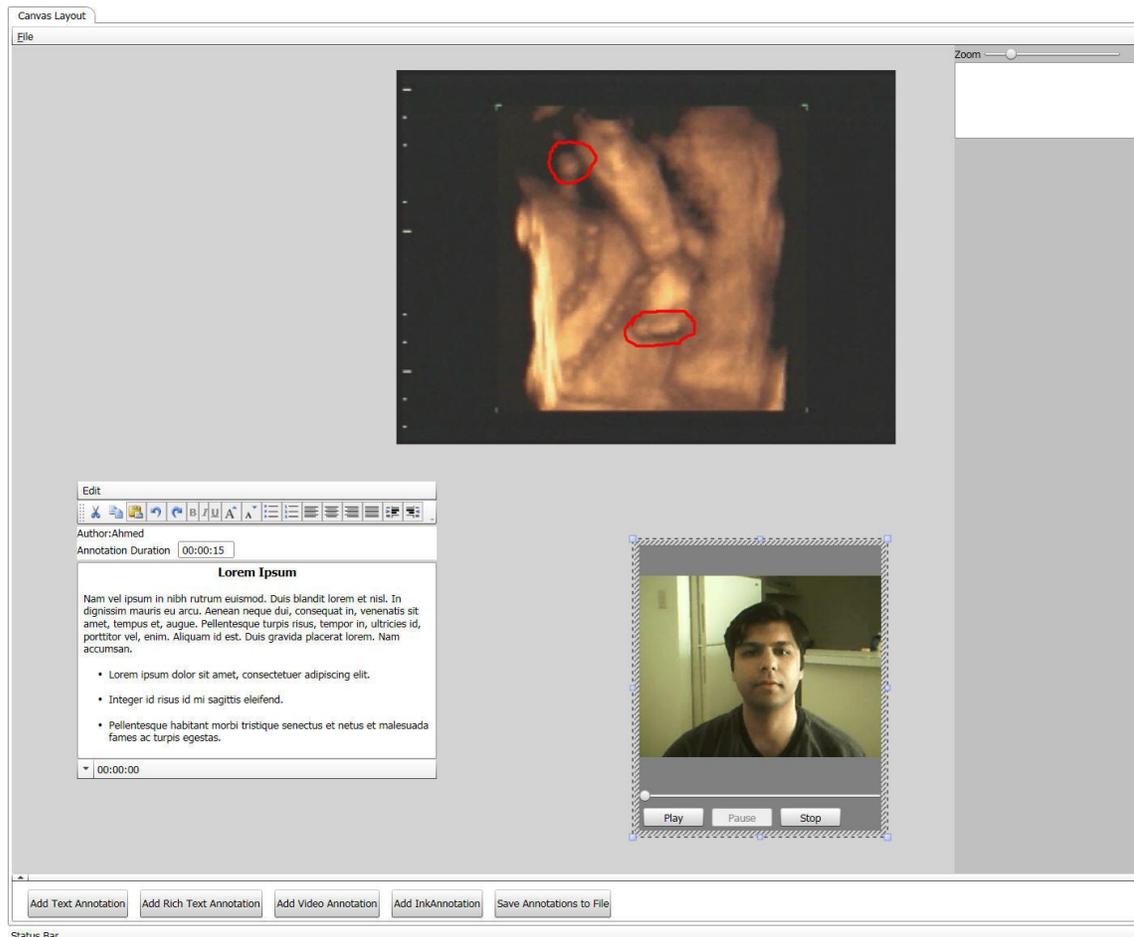


Fig. 8. Annotated Sonogram Video

4.3. SCHOLARLY TEXTS

Scholars are beginning to augment written scholarly work with video based comments to summarize or provide more detail. This is sometimes accomplished with annotating the research paper with video clips of the researcher presenting his work [32]. Another potential use of video clips with respect to scholarly work occurs during the

writing phase of the research. During this phase researchers frequently have to comment on each other's written word. This is mostly done by using document tracking and commenting features of the word processor or by exchanging emails. For some situations this can be cumbersome and a face to face meeting of a few minutes can be much more useful.

For the above two scenarios it would be useful to have the ability to annotate a document with a quick video comment using a webcam. To explore these scenarios and to add this feature to the toolkit a prototype was created.

This prototype allows a user to annotate a text based document with different types of media. The document can be annotated with a text comment, digital ink comment, a video comment recorded with a webcam or a video comment as an existing video file. The prototype supports the Windows Presentation Foundation's (WPF) flow document format. The WPF framework provides support for the text-based and ink-based annotations out of the box but the video annotation was added as part of the current work. There are tools available that allow conversion of Rich Text Format to the WPF flow documents.

The screenshot displays a document titled "Chapter 1" with the following content and annotations:

Chapter 1

Part One

It was a bright cold day in April, and the clocks were striking thirteen. Winston Smith, his chin nuzzled into his breast in an effort to escape the vile wind, slipped quickly through the glass doors of Victory Mansions, though not quickly enough to prevent a swirl of gritty dust from entering along with him.

The hallway smelt of boiled cabbage and old rag mats. At one end stood a large poster of a man's face, which had been there since the war. It was a face of about forty-five, with a heavy black moustache and ruggedly handsome features. Winston made for the stairs. It was no use trying the lift. Even at the best of times it was seldom working, and at present the electric current was cut off during daylight hours. It was part of the economy drive in preparation for Hate Week. The flat was seven flights up, and Winston, who was thirty-nine and had a varicose ulcer above his right ankle, went slowly, resting several times on the way. On each landing, opposite the lift-shaft, the poster with the enormous face gazed from the wall. It was one of those pictures which are so contrived that the eyes follow you about when you move. BIG BROTHER IS WATCHING YOU, the caption beneath it ran.

Part Three

It was a bright cold day in April, and the clocks were striking thirteen. Winston Smith, his chin nuzzled into his breast in an effort to escape the vile wind, slipped quickly through the glass doors of Victory Mansions, though not quickly enough to prevent a swirl of gritty dust from entering along with him.

The hallway smelt of boiled cabbage and old rag mats. At one end stood a large poster of a man's face, which had been there since the war. It was a face of about forty-five, with a heavy black moustache and ruggedly handsome features. Winston made for the stairs. It was no use trying the lift. Even at the best of times it was seldom working, and at present the electric current was cut off during daylight hours. It was part of the economy drive in preparation for Hate Week. The flat was seven flights up, and Winston, who was thirty-nine and had a varicose ulcer above his right ankle, went slowly, resting several times on the way. On each landing, opposite the lift-shaft, the poster with the enormous face gazed from the wall. It was one of those pictures which are so contrived that the eyes follow you about when you move. BIG BROTHER IS WATCHING YOU, the caption beneath it ran.

Annotations:

- Text Annotation:** A green box containing the text "Text Annotation" is positioned over the first paragraph.
- Digital Ink Annotation:** A hand-drawn black star is positioned over the second paragraph.
- Video Annotation:** A video player showing a man in a white shirt sitting at a desk is positioned over the third paragraph.
- Image Annotation:** A book cover titled "Visualizing Information Retrieval Results: A Demonstration of the Ticker Interface" by North A. Brink and Jan D. Pedersen is positioned over the fourth paragraph.

At the bottom of the window, there is a list of timestamps: [11/29/2007 2:05:41 PM], [11/29/2007 2:04:01 PM], [11/29/2007 2:03:40 PM], and [11/29/2007 2:03:23 PM].

Fig. 9. Annotated Document

Figure 9 shows the document annotation prototype. It shows a document with four annotations. The top left annotation is a text based annotation. The top right is a digital ink based annotation and the bottom two are video annotations. The panel at the

very bottom of the figure shows the annotation creation times. It allows the user to navigate through the text. Clicking on these time stamps brings the respective annotated text in view if it is currently not in view.

Two demo videos showing some of the features of the prototype applications are made available with this manuscript along with instructions for viewing them.

4.4. SUMMARY

Creating and evolving the prototype applications resulted in a better understanding of what types of general services are required by a video annotation application. These services were implemented in the form of a toolkit.

The resulting toolkit provides the ability to create video annotation applications and also has limited support for creating applications for annotating text based documents with video. Working with the prototypes also highlighted some shortcomings in the toolkit in its current form. The most notable is the lack of flexibility in storing the annotations. Another area that can be improved upon is making it easier to support new types of annotations. Currently this cannot be done without changing the toolkit internals. More avenues for further development are described in the next section.

5. CONCLUSIONS

This section explains how the toolkit compares to expected results i.e. the feature set described by Grudin and Barger [16] for an ideal general purpose annotation toolkit and provides some direction for further research and development.

5.1. COMPARISON WITH AN IDEAL TOOLKIT

According to Grudin and Barger [16], an ideal toolkit should have the following features:

- Thorough support for common activities
- Extensibility and customizability
- Storage flexibility
- Universal annotation support
- Interoperability among task-specific interfaces

5.1.1. Thorough Support for Common Activities

Thorough support for common activities requires that it should be very easy to add support for common activities like adding, deleting and retrieving annotations to the user interface. The video annotation toolkit in its current form does well in this area. It is very easy to add the ability to create and delete annotations. This ease of use is provided by using the “commands” infrastructure provided by the Windows Presentation foundation (WPF). This allows the developer to associate code with a user interface

element using declarative markup. As part of the toolkit construction WPF commands in the form of methods and classes were created which allows different types of annotations to be created and deleted.

For annotation retrieval the toolkit exposes annotations using a collection. This collection can be queried using the .NET Framework Language Integrated Query system. This allows for sophisticated queries based on simple metadata attributes such as, author, annotation duration, annotation position etc. and are easily written by a software developer using the toolkit.

5.1.2. Extensibility and Customizability

Extensibility and customizability at both the interface and platform levels requires that a toolkit should provide graphical controls whose look and feel can easily be changed and that the toolkit controls be easily substituted by toolkit user developed custom controls. The framework is where the core toolkit functionality lives. At this level, extensibility and customizability requires that the toolkit user should be able to easily extend the toolkit and customize parts of the toolkit (e.g. the persistence mechanism) without affecting rest of the toolkit.

The video annotation toolkit provides extensibility and customization at both levels. At the user interface level it provides “look less” user interface elements whose look and feel can be totally changed simply by providing a different markup without writing any procedural code. This can be accomplished by a designer using a design tool. This facilitates customization at the user interface level. If the toolkit provided user

interface elements do not meet the developer's needs a custom control can be developed that implements interfaces provided by the toolkit. This facilitates extensibility at the user interface level.

At the framework level the developer can provide a custom persistence mechanism for the annotations if the default xml and file system based persistence provided by the toolkit does not meet the application requirements. Similarly the application developer can create a custom video player for advanced scenarios such as the need to provide playback for custom video formats. Both of these scenarios can be enabled by creating custom classes that implement the corresponding interfaces provided by the toolkit.

5.1.3. Storage Flexibility

This is an area where the toolkit is lacking. It only provides one mechanism for persisting annotations i.e. XML and file system based storage. However, the toolkit can be extended to provide other storage mechanisms by implementing provided interfaces as described above.

5.1.4. Universal Annotation Support

Universal annotation support means that an annotation toolkit should allow for annotating any type of media with any other media type. The focus of the video annotation toolkit was to create a toolkit that allows video to be the target for annotations of different media types. However, it also supports annotation of text documents with

video clips, digital ink and text. Annotation of text with text and digital ink is available as part of .Net Framework. Annotation of text with video was implemented as part of the video annotation toolkit.

5.1.5. Interoperability among Task-specific Interfaces

Interoperability among task specific interfaces means that annotations made in one user interface based on the toolkit should be transferable to another user interface based on the toolkit with minimal effort. Since the toolkit provides dedicated classes for storing different annotations and each of these classes provides an XML representation of itself, it is easy to share annotations made by different applications based on the toolkit.

5.2. FUTURE DIRECTIONS

There are a number of features that can be added to the toolkit to increase the number of scenarios where it can be used. Some of these are:

- Enable annotations on other media such as images and 3D models.
- Enable richer on frame annotations like text and images i.e. allow the video frame to be annotated with text, images and shapes. Currently the toolkit only supports digital ink on the video frame itself.
- Enable different mechanisms for storing annotations such as a relational database. Currently the toolkit provides file system based storage and can be extended to support other kinds of storage.

- Enable use of the toolkit in synchronous collaboration scenarios. Currently the toolkit only supports asynchronous collaboration scenarios. It would be interesting to explore how the toolkit can be extended to support synchronous collaboration scenarios.
- Enable easier addition of new types of annotations. Currently this is not possible without changing the internals of the toolkit.

The video annotation prototypes that were built provided some interesting insights into multimedia annotation and resulted in a video annotation toolkit that facilitates annotating a video with different media types and that allows annotation of text documents with video. This toolkit along with toolkits for other research activities like accessing, searching, presenting and analysing video material will move us closer to the development of integrated analysis environments or ‘knowledge studios’ for humanities and other research areas like the ones we have for disciplines such as bioinformatics, mathematics, statistics or engineering [24].

REFERENCES

- [1] D. Bargeron and A. Gupta, "A common annotation framework," Technical Report, <http://research.microsoft.com/research/pubs/view.aspx?type=Technical%20Report&id=524> Accessed: Dec. 2006.
- [2] D. Bargeron, A. Gupta, J. Grudin, and E. Sanocki, "Annotations for streaming video on the web," *CHI '99 Extended Abstracts on Human Factors in Computing Systems* Pittsburgh, Pennsylvania: ACM Press, 1999, pp. 278-279.
- [3] D. Bargeron, A. Gupta, J. Grudin, and E. Sanocki, "Annotations for streaming video on the web: System design and usage studies," *WWW8 / Computer Networks*, vol. 31, pp. 1139-1153, 1999.
- [4] D. Bargeron, A. Gupta, J. Grudin, E. Sanocki, and F. Li, "Asynchronous collaboration around multimedia and its application to on-demand training," *Proceedings of the 34th Annual Hawaii International Conference on System Sciences*, Jan. 2001.
- [5] J. J. Cadiz, A. Gupta, and J. Grudin, "Using web annotations for asynchronous collaboration around documents," *Proceedings of the 2000 ACM Conference on Computer Supported Cooperative Work*, Dec. 2000.
- [6] G. Cherry, J. Fournier, and R. Stevens, "Using a digital video annotation tool to teach dance composition," *Interactive Multimedia Electronic Journal of Computer Enhanced Studies*, <http://imej.wfu.edu/articles/2003/1/01/index.asp>, 2003.
- [7] T. Corrigan, *A Short Guide to Writing about Film*, 5th ed., New York: Longman, 2003.
- [8] P. Donaldson, "XMAS: Cross-Media Annotation System," <http://icampus.mit.edu/projects/xmas.shtml> Accessed: Oct. 2006.

- [9] Dublin Core Metadata Initiative, "Dublin Core Metadata Initiative," <http://dublincore.org/> Accessed: Oct. 2007.
- [10] W. K. Edwards, V. Bellotti, A. K. Dey, and M. W. Newman, "The challenges of user-centered design and evaluation for infrastructure," *Proceedings of the Conference on Human Factors in Computing Systems*, pp. 297-304, Apr. 2003.
- [11] Z. Gang, C. F. Geoffrey, P. Marlon, W. Wenjun, and B. Hasan, "eSports: collaborative and synchronous video annotation system in grid computing environment," *Proceedings of the 7th IEEE International Symposium on Multimedia*, Apr. 2005.
- [12] K. Gocsik, "Writing About Film," <http://www.dartmouth.edu/~writing/materials/student/humanities/film.shtml> Accessed: Nov. 2006.
- [13] D. Goh, *Patron Augmented Digital Libraries*, Ph.D. Dissertation, Texas A&M University, College Station, Texas, 1999.
- [14] D. Goh and J. Leggett, "Patron-augmented digital libraries," *Proceedings of the 5th ACM Conference on Digital Libraries*, Jun. 2000.
- [15] R. Gonzalo and B. Ravin, "Fluid interaction techniques for the control and annotation of digital video," *Proceedings of the 16th Annual ACM Symposium on User Interface Software and Technology*, Nov. 2003.
- [16] J. Grudin and D. Barger, "Multimedia annotation: An unsuccessful tool becomes a successful framework," *Communication and Collaboration Support Systems*, K. Okada, T. Hoshi, and T. Inoue, eds. Tokyo, Japan: Ohmsha, 2005.
- [17] B. L. Harrison and R. M. Baecker, "Designing video annotation and analysis systems," *Proceedings of the Conference on Graphics Interface*, vol. 92, pp. 157-166, May 1992.
- [18] R. Kolker, "Digital media and the analysis of film," *Companion to Digital Humanities*, S. Schreibman, R. Siemens, and J. Unsworth, eds. Malden, Massachusetts, United States: Blackwell Publishing, Incorporated, 2004.

- [19] M. Kornbluh, M. Fegan, and D. Rehberger, "Media matrix: A digital library research tool," *Proceedings of the 5th ACM/IEEE-CS Joint Conference on Digital Libraries*, pp. 412-412, Jun. 2005.
- [20] S. LeeTiernan and J. Grudin, "Fostering engagement in asynchronous learning through collaborative multimedia annotation," *Proc. INTERACT 2001*, pp. 472-479, Jul. 2001.
- [21] S. LeeTiernan and J. Grudin, "Supporting engagement in asynchronous education," *Conference on Human Factors in Computing Systems*, pp. 888-889, Apr. 2003.
- [22] Library of Congress Network Development and MARC Standards Office, "MARC Standards," <http://www.loc.gov/marc/> Accessed: Jan. 2007.
- [23] W. E. Mackay, "EVA: an experimental video annotator for symbolic analysis of video data," *ACM SIGCHI Bulletin*, vol. 21, pp. 68-71, 1989.
- [24] A. Marsden, H. Nock, A. Mackenzie, A. Lindsay, J. Coleman, and G. Kochanski, "ICT tools for searching, annotation and analysis of audiovisual media," Technical Report, <http://www.phon.ox.ac.uk/avtools/index.php> Accessed: Dec. 2006.
- [25] Metavid, "About Metavid," <http://metavid.ucsc.edu/> Accessed: Dec. 2006.
- [26] X. Mu and G. Marchionini, "Enriched video semantic metadata: Authorization, integration, and presentation," *Proceedings of the American Society for Information Science and Technology*, vol. 40, pp. 316-322, Oct. 2003.
- [27] R. Pea, M. Mills, J. Rosen, K. Dauber, W. Effelsberg, and E. Hoffert, "The DIVER™ project: Interactive digital video repurposing," *IEEE Multimedia*, vol. 11, pp. 54-61, 2004.
- [28] S. Pfeiffer, C. Parker, and C. Schremmer, "Annodex: a simple architecture to enable hyperlinking, search & retrieval of time--continuous data on the Web," *Proceedings of the 5th ACM SIGMM International Workshop on Multimedia Information Retrieval*, pp. 87-93, Nov. 2003.

- [29] R. Schroeter, J. Hunter, J. Guerin, I. Khan, and M. Henderson, "A synchronous multimedia annotation system for secure collaboratories," *2nd IEEE International Conference on E-Science and Grid Computing*, Dec. 2006.
- [30] R. Schroeter, J. Hunter, and D. Kosovic, "Vannotea a collaborative video indexing, annotation and discussion system for broadband networks," *K-CAP Workshop on Knowledge Markup and Semantic Annotation*, Oct. 2003.
- [31] R. Schroeter, J. Hunter, and D. Kosovic, "FilmEd-collaborative video indexing, annotation, and discussion tools over broadband networks," *Proceedings of 10th International Multimedia Modeling Conference*, pp. 346-353, Jan. 2004.
- [32] B. Singer, "Hypermedia as a scholarly tool," *Cinema Journal*, vol. 34, pp. 86-91, Spring 1995.
- [33] R. H. Trigg, "Computer support for transcribing recorded activity," *ACM SIGCHI Bulletin*, vol. 21, pp. 72-74, 1989.
- [34] T. Volkmer, J. R. Smith, and A. P. Natsev, "A web-based system for collaborative annotation of large image and video collections: an evaluation and user study," *Proceedings of the 13th Annual ACM International Conference on Multimedia*, pp. 892-901, Nov. 2005.
- [35] K. Weber and A. Poon, "Marquee: A tool for real-time video logging," *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems: Celebrating Interdependence*, pp. 58-64, Apr. 1994.
- [36] Wikipedia, "Microsoft .NET framework 3.0," http://en.wikipedia.org/wiki/.NET_Framework_3.0 Accessed: Nov. 2006.
- [37] Wikipedia, "Windows presentation foundation," http://en.wikipedia.org/wiki/Windows_Presentation_Foundation Accessed: Nov. 2006.
- [38] D. Woods, "Transana (Version 1.22)," <http://www.transana.org> Accessed: Dec. 2006.

VITA

Ahmed Chaudhary graduated with a Bachelor of Science in Electronic Engineering from the Ghulam Ishaq Khan Institute of Engineering Sciences and Technology in May 2003. He is currently a student at Texas A&M University and will graduate in May 2008 with his Master of Science in computer science. He can be reached at the Department of Computer Science, Texas A&M University care of Dr. John Leggett, or you can email him at: ahmedshafi@gmail.com.