Interface Concepts for the Open Video Project

Gary Geisler, Gary Marchionini, Michael Nelson, Richard Spinks, and Meng Yang
School of Information and Library Science
University of North Carolina at Chapel Hill
{geisg, march, mln, spinr, yangm}@ils.unc.edu

Abstract

The Open Video Project is an on-going effort to develop an open source digital video collection that can be used by the information retrieval, digital library, and digital video research communities and ultimately serve an even broader audience. The collection currently contains about 350 video segments from 40 different titles, with a range of characteristics. We believe developing and using this collection as a shared video testbed can help advance progress in video retrieval research by fostering a collaborative research environment. In this paper we first discuss the underlying challenges of compiling and maintaining an open source digital video collection, and then describe the user interface strategies we are developing to enable people to explore and retrieve video from the collection effectively and efficiently.

Introduction

Imagine two undergraduate students studying together for a final exam in a biology course in which the lectures and lab sessions are videotaped and made available as part of the online course materials. In addition to whatever course notes they have taken on their PDAs, they have more than 50 hours of digital video available from the 30 class sessions. How do they browse the video collection to review the course? How do they find specific segments relevant to the topics they need the most help with? How can they share what they've found with other students?

Or, imagine a university researcher whose focus is developing algorithms for automatically recognizing and extracting scenes with faces from video footage. He has developed a new algorithm that he thinks will be effective for several types of situations, such as surveillance videos, news programs, and sports events. How can he obtain enough varied video to rigorously test his algorithm? How can he fairly compare the effectiveness of his algorithm to existing algorithms?

Maybe you have a child who is creating a multimedia project on the topic of hurricanes. Where can she easily find video footage of hurricanes that she can incorporate into her project? Or maybe you're an instructor searching for clips to use in the classroom, a multimedia designer looking through stock video footage, a safety inspector reviewing crash test results, a surgical student needing examples of operating procedures, or a usability researcher managing video records.

We are moving quickly towards a day where these scenarios will be not only realistic but commonplace. Anticipating this reality raises many questions and problems that must be solved to ensure that these scenarios have successful and satisfying outcomes. How do we collect, organize, and store digital video so it can be efficiently found, searched, and downloaded by users with varied interests? How do we help users understand the extent and character of the video content a collection contains, and how can we enable them to search, browse, and preview its contents? How do we administer such a collection so it will continue to grow both in size and scope? Answers to these questions depend as much on human abilities, preferences, and constraints as on technical solutions, and the intersection of these challenges is at the human-computer interface to video repositories.

As a means of exploring these issues, we have developed the Open Video Project, a shared digital video repository and test collection. The primary goal of the Open Video Project is two-fold: to serve as an open source platform for investigating and proposing potential solutions to these and other problems, and ultimately to provide a digital video collection that meets the needs of users in the previous scenarios. Active as a public site at www.open-video.org for more than a year, the collection currently contains about 350 video segments, representing more than 22 hours of footage. This paper first describes more fully the goals of the project and our approaches to developing the infrastructure necessary to support this type of digital video collection, including issues related to metadata, contributions from the video community, and distributed storage. We then go on to describe our efforts to create effective interfaces for browsing, retrieving, and using the contents of the collection.

Project Overview

The Open Video Project was developed at the Interaction Design Laboratory (IDL) at the School of Information and Library Science at the University of North Carolina at Chapel Hill. Building on initial contributions from U.S. government agencies, such as the National Archives and NASA, and from Carnegie Mellon's Informedia project[1], the Open Video collection currently contains several hundred digitized video segments consisting of more than 15GB of footage (primarily in MPEG-1 format). The collection is hosted as one of the first channels of the Internet2 Distributed Storage Infrastructure project[2], which supports distributed collection hosting for research and education in the Internet2 community. To make the video more widely available and to solicit contributions and feedback from the research community, we have published previous papers (Geisler, 2000; Slaughter, et al., 2000) and are currently collaborating with the Open Archives Initiative (OAI)[3] and with the National Institute of Standards and Technology (NIST) Text Retrieval Conference (TREC) team[4].

The type of video currently in the collection consists primarily of government-produced documentary footage (such as "How Water Won the West"), scientific lectures ("Senses and Sensitivity"), historical advertising footage ("Chevrolet Promotional Films"), and educational films from various periods in the 20th Century ("What Makes a Good Party"). To optimize the near-term usefulness of the collection to a wide variety of researchers, we are actively attempting to broaden the contents of the collection, both in size and variety.

Table 1 summarizes the current contents. Our immediate plans are to add other genres of video—including news, entertainment, and even home videos—and to include alternative formats, such as RealVideo, AVI, and QuickTime. The project Web site includes a contribution page (discussed later), and we anticipate that through contributions and our own work, the collection will ultimately reflect a much wider variety of video characteristics in terms of genre, compression schemes, amount of motion, language, sound and color.

[1] http://www.informedia.cs.cmu.edu
[2] http://dsi.internet2.edu/docs.html
[3] http://www.openarchives.org
[4] http://trec.nist.gov

Number of video titles	40
Number of video segments	346
Number of hours	22
Range of segments, in duration	10 seconds to 1 hour
Genre of segments	Documentary, Lecture, Advertising
Color characteristics	275 in color; 71 in black & white
Sound characteristics	305 with sound; 41 silent

Table 1. Summary of video currently in collection

Although other large collections of digital video do exist, such as the Informedia Project (Christel, 1999), the video in these collections is not publicly available. NIST has created a public domain digital video test collection (Schmidt & Over 1999), but it is of limited size and scope. The Internet Archive[5] intends to make public 1000 hours of video, but only a small portion is available as of this writing and the data is only in MPEG-2 format. We are actively collaborating with each of these groups.

By making available video segments with a variety of characteristics, the Open Video collection currently serves as a place where researchers can go to to acquire video to study a wide range of problems, including tests of algorithms for segmentation, summarization and creation of surrogates that describe video content; or interfaces that display result sets from queries. Because researchers attempting to solve similar problems have access to the same video content, the collection can also be used as a test collection that will enable systems to be compared, similar to the way the TREC conference corpuses are used for text retrieval.

Infrastructure for a Digital Video Collection

While doing as much as possible to increase the current value of the collection to the information retrieval, digital library, and digital video research communities, we are also working on a number of fronts to investigate the fundamental problems of creating and maintaining a digital video collection. To set a context for a discussion of this work, consider how much difficulty people have today finding and evaluating text documents from the millions available from the Web. Now imagine how much more difficulty people might have trying to find and evaluate video documents—that have many more features but fewer indexing techniques at this time—when there are millions of those available. If effective user interfaces for interacting with video documents are to be devised, we must first deal with the underlying challenges of compiling and maintaining a digital video collection. Key elements of such an infrastructure include metadata, contributions, and distributed storage. The following sections provide a brief overview of our efforts in these areas.

Content Metadata

Developing a metadata schema for a video collection poses some unique challenges. Most of the metadata research and standards work done to date has been directed at structured and text data, while "practices in the context of digital media and multimedia data management are ad-hoc" (Klas and Sheth, 1994). Thus, the developer of a metadata schema for non-text collections can only look at existing, text-oriented standards as a starting point. For instance, on one level, video can be compared to traditional resources, such as a book, in the type of appropriate metadata that

[5] http://www.openarchives.org

should be collected. The title of the video, year produced, length, director (equivalent to author), production company (equivalent to publisher), subject, keywords, and so on are just as essential high-level attributes as are their text-based equivalents. For the purposes of most general users who are looking for news video on CNN, for example, this level of metadata might be sufficient.

As discussed later, however, much of the expected user audience of the Open Video collection has different, more technical, requirements. These users are more interested in lower-level aspects of video, and there are potentially many such aspects. Video structure is often broken-down into several levels, such as frames (a single still image), clips (a contiguous sequence of frames), shots (one or more clips recorded from the time a camera is turned on until it is stopped), and scenes (a unified collection of shots) (Yeo and Yeung, 1997). Each of these levels could potentially have quite a number of metadata attributes associated with it. Furthermore, unlike text-based collections, the metadata associated with any of these levels could include visual metadata or surrogates, such as thumbnails, storyboards, or even audio excerpts. Table 2 illustrates some of the types of metadata relevant to video content and the issues associated with them.

Type of metadata	**Description**	**Issues**
Descriptive	Bibliographic (title, producer, description, etc.)	In contrast to text, frequently not directly extractable from content
Structural	Size, format, compression scheme, etc.	Can change depending on format; wide range of values
Administrative	Terms and conditions of use, rights management	Copyright issues can be more complex than text (e.g., separate copyrights to audio track)

Table 2. General types of metadata and issues associated with video content

Clearly, for the research community, an additional, more technical level of metadata beyond the descriptive is required. But how much to include is a key decision in creating a metadata schema for the Open Video collection, as the number of potential technical parameters for video is quite large. But with size comes increased complexity and thus increased requirements for administration and maintenance.

Knowing exactly what types of metadata would be useful to researchers is one of the difficulties of creating the Open Video schema. Researchers will be using the content of the collection for many different purposes, and until we fully understand all these purposes, creating a schema that meets the needs of all users will be difficult. We began with a schema containing 30 elements that provided the basic descriptive, structural, and administrative metadata to represent the video, facilitate user interface innovations, and make contributions painless. This schema, however, is expected to evolve as we receive feedback from users.

Community Contributions and Involvement

As the name of the project suggests, the Open Video project is intended to be an open source effort. The content of the collection is freely available (in some cases with research-use restrictions), and we plan to expand the size of the collection in part by encouraging submissions from the digital video community.

This strategy makes a complete and comprehensive metadata schema especially important for the Open Video collection. Members of the community who possess digital video (free of copyright restrictions) that they are willing to share are able to add the video to the collection. Although this helps the collection grow more rapidly, it raises the issue of how to add the metadata from these contributions so that the content is fully integrated into the collection. One method we considered was to ask each contributor to send the metadata to us, which we would then enter into the database in a way consistent with the existing data. This method, however, requires a centralized effort and thus might slow down the growth of the collection, as well as being counter to the open source tradition.

Instead, we have chosen to use a more streamlined method of accepting contributions by requiring the contributor to add the metadata for submissions directly. A Web-based form is available on the Open Video Web site. The contributor enters the data for her submission on the form, submits it, and the metadata is stored in a temporary, mirrored version of the existing metadata tables. The Open Video administrators receive an automatic e-mail notification of each submission. We then check the submission records to ensure they meet the collection requirements and then transfer the records to the production database, making the contributed video available to the end-users.

The use of a controlled vocabulary for the metadata greatly simplifies the process of adding contributed video. The Web-based form provides only defined values for most of the fields (fields like "description" are free text fields), ensuring that submitting video is a relatively quick and easy process. This strategy also ensures that we maintain an internally consistent database, which provides a foundation for effective query and display interfaces.

To facilitate the development of alternate interfaces to the collection by other members of the research community, the Open Video Project will participate as both a service provider and data provider in the Open Archives Initiative (OAI). The OAI is an emerging standard for metadata harvesting and exchange that grew from the scholarly e-print archives community (Van de Sompel & Lagoze, 2000). The purpose of the OAI is to provide a low barrier for digital library interoperability through the harvesting of textual metadata. Metadata harvesting is in contrast to the more difficult area of distributed and parallel searching.

The OAI provides middleware for publishers and builders of digital libraries; OAI compliance and interaction is transparent to users. In OAI terminology, "data providers" build an OAI metadata harvesting protocol compliant layer onto their archive and allow their metadata to be retrieved from robots, either in whole or in part broken out by datestamps or locally defined sets. This allows "service providers" to harvest from multiple data providers, ingest the metadata in their own manner, and provide their own services on the metadata. The metadata is required to have a URI that points back to the originating archive that contains the copy of the resource. Although it is possible for organizations to be both service providers and data providers, the OAI metadata harvesting protocol decouples the two functions and allows data providers to expose their content to multiple service providers, and allows service providers to harvest from multiple data providers.

In the case of the Open Video project, we act as a service provider by providing a number of experimental interface technologies for interacting with our video collection. However, since we also make the Open Video collection available through an OAI metadata harvesting protocol compliant layer, it is possible for others to extract our metadata, possibly even build surrogates on their end, and provide their own alternate interfaces to the videos in our collection, as shown in Figure 1. These alternate service providers could demonstrate different technologies or even filter the content and services to suit the needs of their targeted user communities.

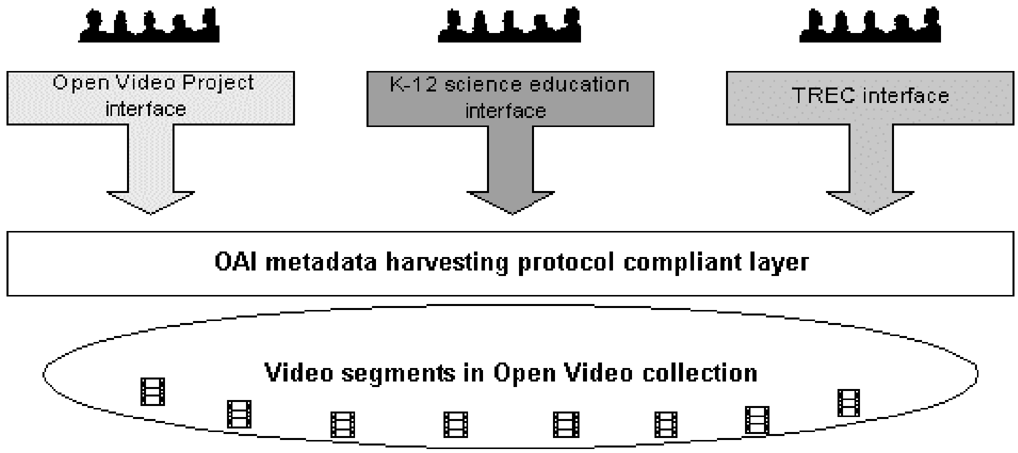

Figure 1. Using an OAI metadata harvesting protocol compliant layer to enable multiple interfaces to the Open Video collection

Finally, the "open" and distributed nature of the Open Video Project provides an on-going case study for researchers interested in the development of an open source community effort, in this case one centered on digital content rather than software.

Distributed Storage and Replication

Digitized video makes for sizeable files, and a digital video collection of the size anticipated by this project necessitates large capacity storage. The Open Video files are currently stored on high-capacity servers provided by the Internet2 Distributed Storage Infrastructure project (I2-DSI). The I2-DSI project is experimenting with architectures for robust, scaleable replication services to enable performance-enhanced delivery of data shared among research and academic partners connected by the Internet2 backbone network. The Open Video Project serves as a prototype "community application" for exploring issues in the I2-DSI project.

In addition, we are experimenting with storing the video in "buckets": aggregative, intelligent publishing constructs (Nelson, 2000). Buckets are aggregative in that they combine data, metadata, and the methods (operations) defined on the data and metadata. This is designed to prevent the information "drift" we have observed in digital libraries when there is a multiplicity of data and metadata types and formats which become "unlinked" and "lost" over time. Buckets are intelligent in that they are entirely self-contained and self-sufficient. They do not depend on the existence (or absence) of any particular DL, repository, database, or search engine. As a result, they manage their own contents, enforce their own terms and conditions, and carry their own source code inside themselves.

Buckets are currently implemented as independent CGI scripts and are accessible through their API with HTTP-encoded messages. The current versions are implemented in Perl. Alternate implementations are being investigated, but the API will remain unmodified. Buckets have been tested on a variety of Unix and Windows NT platforms. Although bucket tools make use of the bucket API, buckets appear as normal Web sites to the casual user.

A bucket is created for each video segment in the Open Video collection. The contents of each bucket can include:

- Metadata
- Video segment (MPEG-1)
- Video segment in other formats / encodings
- Extracted keyframes (GIFs, JPEGs, PPMs, etc.)
- Video abstracts, such as skims, and other video surrogates
- Textual output from "mpeg_stat" analysis tool

The only required elements are the metadata and video segment itself— all other pieces are optional and can vary from bucket to bucket. It is possible to introduce new data elements over time as well. If viewed directly, the output of a bucket appears as:

However, the Open Video interface can extract the metadata from the bucket with:

http://buckets.dsi.internet2.edu/openvideo/buckets/ov-71/?method=metadata

and build direct links to individual elements, such as the MPEG-1 video segment:

http://buckets.dsi.internet2.edu/openvideo/buckets/ov-71/?method=display&pkg_name=video.pkg&element_name=video.mpg

Linking directly to the elements in the bucket enables many interfaces to provide access to bucket contents. Buckets seek to enable advances in interface design by providing a clear separation from the storage granularity of objects and presentation granularity. However, the bucket retains the capability for basic display of its content on its own, should other interfaces become unavailable.

Interfaces for Better Video Browsing and Retrieval

Although we have seen a notable increase in the amount of video material and the number of video collections available in recent years, few of these collections provide their users with interfaces designed for effective browsing and retrieval. Users are primarily dependent on word-based indexing to query and browse. They may be able to type a natural language query that is parsed and matched with manually or automatically-created word indexes, or they may be able to dynamically select from categories that are based on those indexes.

For the Open Video Project we aim to create an interface that provides users with specific features and mechanisms that are optimized to more easily browse and retrieve video. We believe it is especially crucial to provide users with maximum information to inform relevance judgments before accepting the time costs of downloading video. Thus, in addition to the retrieval task, we aim to help people understand a video collection's structure, what is and is not available, and what attributes might be useful for retrieval purposes. We are also providing people with a range of surrogates and integrating these surrogates into an effective and efficient interface. Although some projects have integrated multiple surrogates (for example, Informedia (Wactlar, et. al, 1996) and CueVideo (Ponceleon, et. al., 1999)), we aim to create an environment that provides multiple surrogates at the collection level as well as at the item level.

Exactly how we create surrogates and integrate them into the interface is largely dependent on the needs and characteristics of our user audience. What features of video will best help our users find potentially relevant video segments, and which features of those segments will best help them evaluate whether a given segment is appropriate enough for their purpose to download? Determining this requires an understanding of how humans process video information, the needs of our user audience, and the possibilities for indexing and abstracting video.

Human processing of video information

Theories about how human visual systems process information have evolved over time from Wundt's structuralism (sensory atoms of experience such as color in the visual field are combined and associated with past experiences to cause perceptual images), gestaltism (emergent properties of a whole scene cause perceptual images), Gibson's ecological optics (information in and about the environment striking the retina is the basis of perceptual images), to constructivism (combining retinal information from the environment with emergent properties to construct perceptual images). Many vision scientists study physical features such as color, shape, size, texture, juxtaposition, and motion. Other theorists focus on comprehension of images and how the brain interprets physical information and memories to make sense of scenes. See Palmer, 1999 for an extensive treatment of the different approaches to vision science. Of course, video also contains audio information that is processed by the human auditory system, which in turn has its set of theories. Thus, understanding a video at hand requires humans to integrate multiple information flows, and it is evident that the temporal nature of video and careful crafting of mutually reinforcing messages in the visual and audio channels aids in understanding. Finding the video in a large collection, however, is problematic on several counts.

Unfortunately, the temporal dimension, and often one or more of the information channels, is removed in the surrogates used to find video in repositories. Moreover, we currently have little experience with retrieving video by surrogates other than textual metadata. It is evident that video retrieval is an iterative process that integrates querying, navigation, browsing, and selective viewing. The multivariate features of video make this process more complex and iterative than searching for text files. More cycles of interaction are needed to gain the gist of a video. These inherent complexities are exacerbated in the search process by the fact that digital video files are orders or magnitude larger than text files and people will set higher relevance probability thresholds before downloading primary files for viewing. Thus, we must develop video retrieval systems that provide a variety of surrogates and rapid control mechanisms to move across levels of granularity and search activity.

Diverse audiences

Earlier, we introduced several scenarios where users had a need to access and browse a digital video collection. While the general task was the same, the context surrounding the user's task in each scenario was quite different. This has implications for several aspects of the Open Video collection, including the type of metadata we store, the type of video content we collect, and most important, the design of effective user interfaces to the collection. Before describing the details of our approach to designing a user interface to meets the needs of a diverse user community, we first describe the characteristics of that community in more detail. Generally speaking, we expect users will come from two primary sectors: research and education.

Research community

The research community includes those who are doing some form of work related to digital video and use the content of the collection for their research. This includes researchers working in areas such as information retrieval, digital library, and digital video. They may be studying problems such as creating algorithms for automatically detecting scenes and shots, video abstracting, face or feature detection, summarization and creation of surrogates that describe video content; or developing interfaces that display result sets from queries of multimedia material. These sorts of research questions involve examining a variety of different aspects of video clips, such as color characteristics, pixel changes from frame to frame, and qualities of the audio track.

Education community

The education community could potentially be a significant user of the Open Video collection. Video is increasingly being created and used in traditional classroom settings as well as in the blossoming area of distance education. Instructors may want to search the collection for content they can use for teaching or supplementing instruction on particular topics. Students, too, may look to the collection for material they can use to supplement their learning of particular topics or to integrate into multimedia projects. The current content of the collection has important potential for science and social studies education.

In addition to the relatively focused purposes of the research and education communities, the Open Video collection may become useful in a broader way to the general public. Schoolchildren might explore the collection to find clips for projects, multimedia artists might browse to find content to integrate into their work, and the proliferation of digital video cameras may encourage people to contribute video.

Because the initial users of the collection are from the research community, however, we understand their needs best at this stage. The interface prototypes we describe below therefore focus on users from the research community. As work on the Open Video project evolves, however, we are particularly interested in developing solutions that are flexible enough to meet the needs of each of these user communities.

Indexing and Abstracting for Retrieval

Retrieval depends on good indexing and abstracting and the user's ability to articulate queries that take advantage of this indexing and abstracting. However, even after forty years of information retrieval progress, search remains a challenging function in electronic text environments. Video retrieval will surely not be an easier task for people. The temporal nature and multiple channels of video content emphasize the need for search and browsing mechanisms that offer more representation facets than text. In the sections below we briefly summarize issues related to video indexing and abstracting and then describe how we will incorporate research in these areas into a user interface for browsing and retrieving video.

Video Indexing

Indexing video requires some form of content-based analysis of the video features. Video features include both visual and audio features. Tables 3 and 4 present a brief view of the visual and audio features of video.

Visual Feature	**Examples**
Color	hue, chromo, saturation, line frequency, line lengths, angle frequency, edge magnitude
Shape	area, circularity, eccentricity, major axis orientation
Texture	coarseness, contrast, periodicity, randomness, directionality
Text	captions, credit titles
Faces	
Motion	camera motion: panning, tilting, zooming, tracking, booming and dollying
Gradual effects	dissolving, fading in, fading out, matte, translates

Table 3. Visual Features

Audio Feature	**Examples**
Physical properties	amplitude, frequency, waveform
Psycho-acoustical properties	speech, music, silence, noise, violence

Table 4. Audio Features

Researchers have developed many techniques for feature detection and extraction such as cut detection, motion analysis, and face recognition. Many projects use a combination of features to index the video. The QBIC (Query by Image and Content) project (Ashley et. al., 1995) indexes and retrieves images according to average color, histogram color, texture or pattern, shape and position. The JACOB system (Ardizzone, 1996) develops automatic video indexing by color, motion and shape information. Zhang et al. (1995) proposed an automatic image-indexing tree, which mainly uses texture and color. Additionally, besides the visual elements of the video, the text information in the video such as the captions and titles can provide more direct and easy understanding of the information of the video. Current Optical Character Recognition (OCR) technology can be used to do the text extraction. The Video OCR project (Sato, 1999) proposed techniques for character recognition for news video. Researchers at the University of Mannheim (Lienhart, 1997) presented algorithms for automatic character segmentation in motion pictures, as well as for automatic extraction and identification of text in pre-title sequences, credit titles, and closing sequences with title and credits.

Video Abstracting

Publishers and librarians have long created surrogates (for example, identifiers, abstracts, and synopses) for information objects so that these alternative representations could aid retrieval. In browsing, surrogates provide an important alternative to primary objects because they take far less time to examine and provide enough semantic cues to extract gist and allow users to assess the need for further processing of other surrogates and the primary object.

Video surrogates can be classified into two types: still-image and moving-image. Still-image abstracts are easier to obtain and thus more commonly used. Slide shows, storyboards, and filmstrips are examples of this type (Christel, 1998). Usually, people extract keyframes from each shot to represent it and then arrange all or a subset of the keyframes to form the abstract. The methods of selecting the keyframe and of clustering or assembling the keyframes vary in different projects. For instance, within the Informedia collection, the shot's middle frame is assigned by default to be the keyframe. Marchionini et al. (1997) used the first "I" frame after a scene change as the keyframe. In the Vabstract project (Pfeiffer, 1996), high-motion scenes and high-contrast scenes are extracted to represent the feature films.

A moving-image abstract is a short video itself and can provide rich and vivid information for the users. The most recognizable example is the movie trailer. The video skim used in the Informedia project is also such a type (Christel, 1998). The Movie Content Analysis (MoCA) Project (Lienhart, 1997) selects some clips of a film and then assembles them into the final abstract. Moving-image abstracts incorporate both audio and visual information from a longer source and can be regarded as a short preview for a longer video.

As more digital video becomes available and digital video retrieval research progresses, there surely will be a multiplicity of surrogates—both still-image and moving-image types—created. There has, however, been very little work on studying how people interact with and use such surrogates. For example, Eliot (1993), Mills et al. (1992), and Tonomura et al. (1993) each proposed innovative user interfaces that respectively included video streams made up of slices of each frame; hierarchical thumbnails; and multiple surrogates for features such as brightness, "x-ray" compositions, and color. However, none of these tools was tested with users.

The Informedia Project is an important exception, having both constructed innovative interfaces for video retrieval and conducted studies of how people use alternative surrogate implementations (Christel et al., 1998; Christel et al., 1997; Smith & Kanade, 1998). Marchionini and his colleagues have conducted user studies to identify key parameters for video browsing (Ding et al., 1997; Ding et al., 1999; Komlodi & Marchionini, 1998; Slaughter et al., 1997, Tse et al., 1998) and have developed a set of methodologies that may be used in other video browsing efforts. However, these studies mainly focused on surrogates for specific video objects and item recall types of tasks. They represent a base upon which to build a video browsing theory and systems that cover more general tasks, user communities, and level of browsing.

We aim to understand broader issues such as how these surrogates are used to extract meaning and to make ongoing decisions that guide browsing within retrieval tasks. More generally, we are concerned with understanding which surrogates are useful for which types of tasks and how people are able to incorporate these surrogates into their information processing strategies.

A User Interface Based on AgileViews

The Web exemplifies the recognition that people must be able to browse content as well as execute queries. In developing a user interface to the Open Video collection, we have tried to take into account both the needs of our user communities and the unique abstracting and indexing features of video. Our goal is to enable an individual user to easily search and browse the collection and evaluate the contents of the collection in a way that makes sense for their particular needs. More specifically, we want to provide users with:

- Overviews of the collection
- Search results by video title or by individual video segment
- Ways to smoothly transition from collection-level browsing to segment level browsing
- A range of quick, effective previews of segments, appropriate to different user goals
- Smooth and convenient access to supplementary information the user can use to make browsing and retrieval decisions

To develop an interface that meets these requirements, we have based our work on a user interface framework developed by the authors called AgileViews (Marchionini et al. 2000). The AgileViews framework attempts to improve information seeking by providing users with agile control over information spaces through the use of alternative views and simple mechanisms for coordinating these views. By explicitly defining overviews, previews, peripheral, and shared views, we believe we can create an interface that will help a diverse range of users effectively access the video content in the Open Video collection.

Overviews and Previews

The primary object of interest in the Open Video collection is the *video segment*. One or more segments make up a *video title.* All available video titles make up the entire *collection*. When an interface provides a user with a summary or view of an entire collection, that is considered an overview of the collection. One way to provide an overview of a collection is to simply summarize its contents, as is done in Figure 2.

Contents as of 1/24/2001		
Video titles:	40 different video titles	
Video segments:	195 individual video segments	
Color:	124 in color	71 in black & white
Sound:	154 with sound	40 silent
Segment range, file size:	0.396 MB to 475.0 MB	
Total size of repository:	12861.574 MB	

Figure 2. Overview of the Open Video collection [not up to-date]

While this overview does provide the user with some useful information about the contents of the collection, and might be appropriate for a collection of text documents, for a video collection it fails to convey any sense of what the content "looks like". Taking into account the unique characteristics of video and the research-oriented needs of our initial set of users, we might improve our overview by providing a visual summary of the collection, indexed by the metadata most relevant to our users. As Figure 3 shows, an overview of the collection, organized by genre, not only gives the user some visual information about the titles in the collection, but by showing each video title displayed on a separate tabbed panel according to which genre it is identified with, the user is able to focus on the part of the collection that is most appropriate for his or her needs.

Figure 3. Overview of the Open Video collection by genre

In addition, by selecting one or more thumbnails from the overview, the user can quickly build a results list to be examined further, as shown in Figure 4. This example also demonstrates how users can move from high-level overviews of the entire collection to previews of specific segments (discussed next) by manipulating a variety of indexes that describe visual and audio features such as human faces, colors, motions, and sounds, as well as words.

Preview	Video Title	Video Segm...	Download	File Size	Duration
	Hurricanes	Segment 1		45.824	234.97
	Hurricanes	Segment 2		31.360	160.49
	New Indians	Segment 1		9.984	56.29
	New Indians	Segment 2		4.992	28.19
	Britton, South Dakota	Segment 1		62.200	279.02

Figure 4. Search results from a digital video collection

Consider two researchers, each searching the video collection for a specific type of segment. The first researcher is working on a face recognition algorithm and wants to find a segment that contains an example that involves human interaction. Moving the mouse over the small preview icon in the search results for a segment displays a pop-up window that contains an array of keyframes from that segment. In Figure 5, this array of keyframes is indexed by faces, which helps the first researcher quickly evaluate the likelihood that the segment will contain an example that meets his needs.

Figure 5. Array of keyframes, indexed by Faces

The second researcher, on the other hand, might be interested in segments that contain high-contrast scenes, such as those taken outdoors. This researcher can display the pop-up window of keyframes for the same segment, but in this case would select the Brightness tab to immediately view an array of keyframes indexed by overall brightness, thus displaying scenes likely to occur outdoors, as shown in Figure 6. The combination of easily accessible pop-up windows with content indexed by useful features enables users with different needs to more quickly evaluate whether available resources will meet their specific needs.

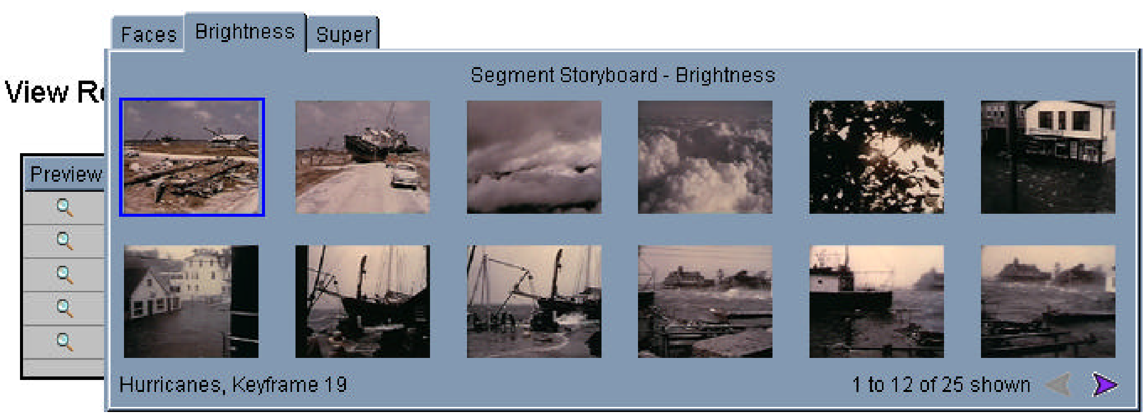

Figure 6. Array of keyframes, indexed by Brightness

History, Peripheral, and Shared Views

Overviews and previews are the most obvious views that people will find useful in exploring and retrieving information. However, histories (reviews), peripheral views, and shared views will also be investigated. The simplest history is to back up through the actions already taken. For histories across sessions, bookmark functions and client-based logs are built in at the browser level. Simple additions such as highlighting will be added in the AgileViews environment to indicate that video objects have been downloaded, annotated, or previewed. More interesting challenges arise with peripheral views (contextual displays and mechanisms) and shared views (real time coordination, either conscious or not).

An example of a peripheral view in the Open Video context might be information about the previous history of use for a given video title or segment. For example, consider the education-based scenario introduced earlier, in which two students need to browse the collection to study for a final exam. A graph or other display that shows how frequently segments have been accessed during the semester may provide useful clues to the students as to which segments others have found useful. This form of peripheral view would become even more valuable if students and the instructor agreed to characterize themselves and include those characterizations as part of the peripheral information available for students registered in the course. More sophisticated community-based models will be monitored and considered as peripheral views for the Open Video collection (for example, the ibiblio project at UNC is investigating the group trust metrics[6]).

Shared views pose an interesting challenge for general use but have an obvious analog in the educational scenario. As two students study together, a picture within picture technique or parallel windows such as those common in collaborative work environments seems like an obvious way to coordinate search. A less obviously useful surrogate would be to have a graph or other display of the current system status available. For example, the peripheral view described above that shows frequency of access across the semester could be rendered as a real-time display of who is using the system for what video at the present time. Although it is unclear how useful this might be for study purposes, it is reasonable to believe that knowing what the rest of the class is studying contemporaneously would be of interest to some students. If collaborative patterns were encouraged by the instructor, e-mail or class listserv services could be leveraged to make studying a more cooperative experience, and more learners could benefit from the serendipitous discoveries and insights made by individuals. We recognize the speculation inherent in these examples, but believe that they are worth exploring by providing the capabilities in the AgileViews environment.

Future Directions

We are actively pursuing several threads as we move forward in developing the Open Video project. In addition to continuing to grow the size and scope of the collection and working on creating effective interfaces for searching, browsing, and retrieving video, we are interested in using the collection to investigate issues related to evaluation, both in the sense of serving as a generalized test collection, and as a specific instance to develop evaluation metrics for video—as opposed to text—collections.

Open Video as a Test Collection

The importance of test collections is best demonstrated by the TREC corpus coordinated by NIST. TREC provides a large text corpus from different sources and sponsors an ongoing set of studies and conferences that advances the state of retrieval engine research as well as the methodologies of information retrieval in general (Harman, 1993; Voorhees & Harman, 1998). The TREC 2001 conference will for the first time include a Video track, designed to investigate content-based retrieval of digital video. Some of the video content to be used for this track was obtained from the Open Video collection.

While the inclusion of Open Video content at TREC is a good indicator of its potential value in serving as a test collection, we want to stress its much broader potential in this regard. To begin any type of video retrieval project requires a great initial start-up effort to locate and digitize appropriate video. Researchers are often able to get donations of video from commercial television and the film industry but then are bound to agreements that prohibit those materials from being distributed outside the project. This is one explanation of why it is currently difficult

[6] http://www.ibiblio.org

to replicate or make comparisons in video retrieval research; since it would not be easy to share materials used for testing with others. By providing a large, varied collection of content in the public domain, the Open Video project can help researchers begin projects more quickly and accurately compare results with others doing similar work.

The size and "open" nature of the Open Video collection can also be valuable in helping explore the properties that will make a video test collection truly useful. For example, some collections used for testing are so small that they cannot be considered a representative sample. The size of the Open Video collection should partially alleviate this problem, but what other properties make it a valuable collection for researchers in a range of disciplines? The open-source nature of the collection makes it ideal for the research community to collectively work towards an effective design of the contents that avoids some of the problems encountered with other test collections.

Evaluation

The lack of previous examples of creating a large collection of video intended for a varied audience raises several questions regarding evaluation. How can we measure what it means to be useful, easy, and effective in a digital video environment? In information retrieval studies, recall and precision values are used as performance metrics to assess the final products of search. But can these be usefully applied to video, which has very different characteristics from text? We believe that a new set of instantiations of these constructs appropriate to the tasks and the media is needed for video retrieval research.

As indicated earlier, the characteristics of video require different forms of surrogates and distinct methods of searching and browsing. Evaluating browsing behavior is a challenge in any medium and we will build upon a set of tasks and metrics developed in our previous work to go beyond traditional metrics to address time-benefit tradeoff measures. We will conduct three types of user studies:

- **Test the usability of a particular mechanism:** When we have designed and prototyped an individual interface mechanism or widget, we will test the usability of that mechanism. This will help us identify any usability problems associated with a particular mechanism, which can then be resolved and the revised prototype can again be tested.

- **Test the usability of combinations of mechanisms:** Once a suite of individual mechanisms for video browsing has been implemented to an acceptable level of usability, they will be grouped together in the AgileViews architecture. Even if each mechanism is quite usable, it does not necessarily follow that the suite of mechanisms will be usable. For example, they may contain internal inconsistencies that are confusing to potential users, or they may leave gaps in functionality that need to be covered for the set to be considered complete. Thus, the graceful integration of the set of implemented mechanisms will be evaluated.

- **Test the usefulness of combinations of mechanisms:** Once we are confident that people can use the suites of mechanisms implemented in the project, we will be in a position to evaluate their usefulness in the "real world." We will address usefulness on two levels. First, we are concerned with the time-benefit tradeoffs offered by different video surrogates and interface mechanisms and how time considerations affect browsing behavior. Second, we seek qualitative assessments of browsing behavior using the AgileViews environment as a summative evaluation.

Usability testing will be done at each iteration of system development and will depend on data at several levels of specificity, ranging from fine-grained eye movements to verbal self-reports. From the results of these studies, and from the results of other digital video projects, we believe we can begin to devise new metrics that will be useful in evaluating video-related systems.

Conclusion

Prominent amid the rapidly expanding wealth of information available is the increasing volume of digital video content. Given the difficulty users have finding relevant text-based information today, we can only expect that these difficulties will be compounded by the added complexities of video content. With the significant amount of research being done today in many areas related to digital video, there is clearly a need for a large-scale, easily accessible collection of digital video that can be used by people with a variety of needs.

Through the Open Video Project we are creating a digital library that we hope will serve these needs and much more. We are attempting to create an environment that allows people to explore and retrieve digital video effectively and efficiently. Creating and studying this environment will also advance progress in video retrieval research by fostering a collaborative research environment. To date we have assembled a substantial sample of open source digital video, developed a distributed storage and delivery architecture that provides for user community contributions, investigated different metadata schemes, and developed several prototype interfaces. The current interface in the operational system offers some of the features we are designing, allowing people to look at the entire collection or different previews for specific segments. This paper illustrates prototype designs for the next iteration of an agile views environment and our work toward evaluating these designs.

References

Ardizzone, E., La Casia, M., Di Ges'u, V., and Valenti, C. (1996). Content-based indexing of image and video databases by global and shape features. In Proc. of the International Conference on Pattern Recognition, 140-144.

Ashley, J., Barber, R., Flicker, M., Hafner, J., Lee, D., Niblack, W., and Petkovic, D. (1995) Automatic and semi-automatic methods for image annotation and retrieval in QBIC *Proceedings of SPIE*. 2420, 24-35.

Christel M., Hauptmann A., Warmack A., Crosby S. (1999). Adjustable Filmstrips and Skims as Abstractions for a Digital Video Library. *IEEE Advances in Digital Libraries Conference, (Baltimore, MD, May)*, 19-21.

Christel, M., Smith, M., Taylor, C.R., & Winkler, D. (1998). Evolving video skims into useful multimedia abstractions. Proceedings of CHI '98: Human Factors in Computing Systems (Los Angeles, April 18-23, 1998), 171-178.

Ding, W., Marchionini, G., & Soergel, D. (1999). Multimodal surrogates for video browsing. Proceedings of Digital Libraries ’99. the fourth annual ACM conference on digital libraries (Berkeley, CA, August 11-14, 1999), 85-93.

Ding, W., Marchionini, G., & Tse. T. (1997). Previewing video data: Browsing keyframes at high rates using a video slide show interface. Proceedings of the International Symposium on Research, Development, and Practice in Digital Libraries, (Tsukuba, Japan), 151-158.

Geisler, G., and Marchionini, G. (2000). The Open Video Project: A Research-Oriented Digital Video Repository [short paper]. In Digital Libraries '00: The Fifth ACM Conference on Digital Libraries (June 2-7 2000, San Antonio, TX), 258-259.

Harman, D. (1993). The DARPA TIPSTER project. SIGIR Forum 26(2), 26–28.

Klas, W., & Sheth, A. (1994). Metadata for Digital Media: Introduction To the Special Issue. SIGMOD Record, 23(4).

Komlodi, A. & Marchionini, G. (1998). Keyframe preview techniques for video browsing. Proceedings of ACM DL '98. (Pittsburgh, PA, June 24-26, 1998).

Lienhart , R., Pfeiffer, S., Effelsberg, W. Video Abstracting. *Communications of ACM*, 40,12(Dec. 1997), 54-62.

Lienhart R. and Stuber F. (1996) Automatic text recognition in digital videos. Proceedings of the SPIE Image and Video Processing IV. 2666: 180-188.

Marchionini, G., Nolet, V., Williams, H., Ding, W., Beale, J., Rose, A., Gordon, A., Enomoto, E., & Harbinson, L. (1997). Content+Connectivity=Community: Digital resources for a learning community. Proceedings of ACM DL '97. (Pittsburgh, PA, July 23-26, 1997), 212-220.

Marchionini, G., Geisler, G., and Brunk, B. (2000). AgileViews: A Human-Centered Framework for Interfaces to Information Spaces. Proceedings of the Annual Conference of the American Society for Information Science, November, 2000, 271-280.

Nelson, M. L. (2000). Buckets: Smart Objects for Digital Libraries. Ph.D. Dissertation, Old Dominion University, Department of Computer Science. http://mln.larc.nasa.gov/~mln/phd/

Palmer, S. (1999). Vision science: Photons to phenomenology. Cambridge, MA: MIT Press.

Pfeiffer, S., Lienhart, R., Fischer, S., and Effelsberg, W. (1996). Abstracting Digital Moves Automatically. *Journal of Visual Communication and Image Representation,* Vol. 7(4), 345-353.

Pfeiffer S., Fischer S., and Effelsberg W. (1996). Automatic Audio Content Analysis *ACM Multimedia 96,* Boston MA USA.

Ponceleon, D., Amir, A., Srinivasan, S., Syeda-Mahmood, T., & Petkovic, D. (1999). CueVideo: Automated multimedia indexing and retrieval. ACM Multimedia '99 (Orlando, FL, Oct. 1999), 199.

Sato, T., Smith, M., Hughes, E. Kanade, T. (1999). Video OCR: Indexing digital news libraries by recognition of superimposed caption. *ACM Multimedia Systems Special Issue on Video Libraries*, 7 (5): 385-395.

Schmidt, C. & Over, P. (1999). Digital Video Test Collection. In proceedings of the Twenty-Second Annual International ACM SIGIR Conference, Berkeley, California, USA.

Slaughter, L., Marchionini, G. and Geisler, G. (2000). Open Video: A Framework for a Test Collection. Journal of Network and Computer Applications, Vol. 23(3). San Diego: Academic Press, 219-245.

Slaughter, L., B. Shneiderman, and G. Marchionini. (1997). Comprehension and object recognition capabilities for presentations of simultaneous video keyframe surrogates. In Peters C. and C. Thanos (Eds.), Research and Advanced Technology for Digital Libraries: Proceedings of the First European Conference, (pp. 41-54). ECDL'97, Pisa, Italy.

Smith, M. & Kanade, T. (1998). Video skimming and characterization through the combination of image and language understanding. Proceedings of 1998 IEEE International Workshop on Content-based Access of Image and Video Database. (Bombay, India, January 3, 1998). Los Alimitos, CA: IEEE. 61-70.

Tse, T., Marchionini, G., Ding, W., Slaughter, L.& Komlodi, A. (1998). Dynamic keyframe presentation techniques for augmenting video browsing. Proceedings of AVI '98: Advanced Visual Interfaces (L' Aquila, Italy, May 25-27), 185-194.

Van de Sompel, H., and Lagoze, C. (2000). "The Santa Fe Convention of the Open Archives Initiative", D-Lib Magazine, 6(2). http://www.dlib.org/dlib/february00/vandesompel-oai/02vandesompel-oai.html

Voorhees, E. & Harman, D. (eds.) (1998). The Sixth Text Retrieval Conference (TREC-6). NIST Special Publication 500-240. Gaithersburg, MD.

Wactlar, H. D., Kanade, T., Smith, M. A., & Stevens, S. M. (1996). Intelligent Access to Digital Video: Informedia Project. Computer.

Yeo, B.-L., & Yeung, M. M. (1997). Retrieving and Visualizing Video. Communications of the ACM, 40(12), 43-52.

Zhang, H., Zhong, D. (1995) A scheme for visual feature based image indexing *Proceedings of SPIE*. 2420, 36-46.