Journal of Network and Computer Applications (2000) **23**, 000–000
doi:10.1006/jnca.2000.0112, available online at http://www.idealibrary.com on IDEAL®

Open video: A framework for a test collection

Laura Slaughter, Gary Marchionini and Gary Geisler

School of Information and Library Science, University of North Carolina at Chapel Hill, USA. E-mail: (lauras,march,geisg)@ils.unc.edu

The future will bring widespread access to large digital libraries of video. Consequently, a great deal of research is focused on methods of browsing and retrieving digital video. This type of work requires that investigators acquire and digitize video for their studies since the video information retrieval community does not yet have a collection of video for research purposes. There is a clear need for a shared test collection that they can use. This paper provides a framework for such a test collection and describes the Open Video Project that has begun to develop a test collection based on this framework. The proposed test collection is meant to be used to study a wide range of problems, such as tests of algorithms for creating surrogates for video content or interfaces that display result sets from queries. An important challenge in developing such a collection is storing and distributing video objects. This paper is meant to layout video management issues that may influence distributed storage solutions. More specifically, this paper describes the first phase for creating the test collection, sets guidelines for building the collection, and serves as a basis for discussion to inform subsequent phases and invite research community involvement.

1. Introduction

It is inevitable that the technical limitations that impede widespread usage of video libraries will diminish in the years ahead. In preparation for a future where video becomes a ubiquitous information source, there is much research currently underway that is focused on solving problems of retrieval from digital video collections. For instance, investigators have explored automated segmenting of video [1–4] and use of video surrogates [5–11] to improve user access to video data within large video libraries. In order to test new algorithms, search strategies, and interface designs that facilitate browsing and searching video data, researchers must gather and digitize their own collections of video. These researchers each build their own video test collections that are then employed to answer research questions.

Video test collections are generally created by researchers through agreements with organizations that allow the use of their video materials for specific research projects. For a research community committed to investigating video retrieval questions, there are advantages to collectively building a large video collection for use as a community test collection. The creation of a test collection where the contents are well documented and designed according to a formal plan paves the way for researchers to define sets of queries and work with a recognized set of

1084–8045/00/000 000+00 $35.00/0

video documents. Additionally, the ability to replicate experiments and to make comparisons between various techniques will be realized. This paper presents a framework for such a video test collection. The Open Video test collection that is proposed will be used to study a wide range of questions. For example, questions relating to interfaces for browsing video, types of user queries applied to video, or different user browsing strategies. Although frameworks for the 'ideal' test collection have been written for text experimental collections, it is only recently that researchers have had the capability to realistically concentrate on video and other multimedia as an extension of this work.

These new capabilities are dependent on many information technology advances but offer special challenges to the distributed storage (DS) community. In particular, the movement of large files with associated metadata, test protocols, and possibly user interfaces among researchers distributed around the world will benefit from DS techniques that move nodes physically closer to the user. The Open Video project described here benefits from participation in the I2-DSI project described in this issue. The Open Video collection is a channel in the I2-DSI infrastructure.

1.1 *Background*

For users of video collections that are stored in tape format, the process of finding needed video segments can be time consuming and expensive. Imagine a teacher in need of a short segment on water pollution or a documentary film production crew looking for a specific 10-second clip of an event. After identifying the titles of several possible full-length videos that may or may not contain the needed clips, it is necessary to view the entire contents of each before selecting the desired video segments. Searching by title or broad topics to locate segments of video may require that the user view hours of videotape. This can be a tedious process, causing video resources to be used less often than they would be if specific content chunks were easier to find. Video can be an effective and practical information source, but if the user is unable to locate what they are searching for it will be underutilized.

Digital video collections make retrieval easier and offer the potential of allowing users to obtain the maximum benefit from video data. Instead of searching sequentially through tape, random access of any part of the video is possible when it is in digital format. This eliminates the frustrating wait to get to a known segment of video and also permits users to jump ahead at various intervals when scanning for needed content. Moreover, digital video opens up new possibilities for indexing and abstracting that would be very difficult to achieve otherwise. Just as text retrieval has benefited from computational strategies such as statistical and probabilistic analysis of word/phrase occurrences, so video retrieval will benefit from computational analysis of visual elements. What elements are useful and how to process them are important research issues.

Automatically created salient images and short clips to describe the essence of the entire video are two examples of how digital video can help users to rapidly select relevant clips. The ability to find needed video quickly is gained from the storage of video in a digital format and insures greater use and reuse of the video data.

The amount of video data currently available in digital format is continually increasing—especially as digital cameras and personal computing editing tools become more commonplace. Thus, large collections of digital video are no longer just a dream for the future. The types of video data that could be included in such collections may consist of any of numerous genres. Genres for video include news programs, television shows, advertisements, feature films, documentaries, government sponsored programs, surveillance video, and so on. The existence of digital video collections will make it easier for users to locate desired segments of video, get access to the video, and manipulate video data. As storage costs decline in comparison to networking costs, video repositories will depend on creative ways to manage bandwidth-storage tradeoffs.

In order to get information from a large digital collection, a system for searching and retrieving the contained documents must be developed. The idea of searching in electronic environments for text documents has been studied extensively though the use of test collections [12]. The importance of testbeds is best demonstrated by the Text Retrieval Conference (TREC) corpus coordinated by the National Institute of Standards and Technology (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 [13,14].

Technology has advanced so that it is now necessary to extend our knowledge to digital video documents. This includes learning how to segment video into meaningful chunks and how to make effective surrogates for video. For instance, natural story units within video need to be recovered from the internal content of the video file. The ending and beginning of a unit of content is not always correlated with a scene change, which is the most currently used method for detecting the ‘edge’ of a segment. The characteristics of video introduce a new set of indexing and searching problems. Research questions will focus on exploration of the different ways that users will want to query video, how they make use of video surrogates and what strategies they use for browsing video. Many researchers are currently working on these issues and, just as it was for text document retrieval research, they have built various test collections to help them answer their questions. Moreover, the National Institute of Standards and Technology has recognized this need and has begun developing TREC-like corpus for video [15].

1.2 *Video retrieval research*

A small sample of projects are discussed in order to illustrate the variety of research problems the video retrieval community addresses and the diversity of

test collections that they create. Even though there are other defining details that describe each project, the major focus of this section is to briefly outline each project in terms of the types of questions under investigation and the test collections used to answer those questions.

Chua and Ruan [16] from the University of Singapore have worked on a frame-based video retrieval and sequencing system (VRSS). The overall goal of their project is to study tools to manage video data efficiently. There were a number of types of tools built. Two examples of these are a semi-automated method for dividing video into meaningful shots and a scene editor that allows users to group sets of shots into a higher-level entity. In order to test their retrieval techniques, they used a single 16-minute video documentary about the National University of Singapore. This video was segmented into 190 video shots and 52 video scenes. The text and cinematic properties of the video were manually entered. Nine queries were formulated for testing purposes and for each query a set of relevant shots was predetermined. The queries were designed to evaluate specific aspects of the system. Unfortunately, no information was given about the quality of the video used. It is unknown whether the video was professionally produced and if there was a written script for action.

Zhang *et al*. [17] also from the University of Singapore, automatically parsed and extracted keyframes within video for indexing purposes. These were used in a system for content-based retrieval based on temporal and keyframe features. The system performs 'queries by example' such as finding similar shots based on camera panning motions, average colours, and colour histograms. They used two 'stock footage' raw unedited videos of various lengths, an elaborately produced documentary, and a dance video for their experiments. The total length of all video footage combined was 66 minutes. These were chosen because the authors felt that the videos represented both 'the content material and style of presentation that one would generally find in professionally produced television and film'.

Wolf *et al*. [18] constructed a video library for use over the World Wide Web. The library was an experimental test collection for two types of algorithms that support access to multimedia material. These were browsing over individual programs and searching over the entire collection. Annotations and keyframe surrogates were included in the collection. Users textually searched the collection in order to narrow down a result set. The library applied the use of keyframes for browsing videos in the result set by arranging them on storyboards and in a slide show format. This test collection was designed so that it would be useful in realistic situations with an audience in K-12 education. The collection contained news reports, political advertisements, NASA documentaries and selected videos from the Presidential Libraries. Much of this video is in the public domain and all titles were chosen because of their historical significance. There were over 40 titles, all in MPEG-1 format, ranging from 30 seconds to 1.5 hours in length. The videos were segmented into small clips for teachers to use in their classrooms.

Gauch *et al*. [19] created the VISION digital video library (Video Indexing for Searching Over Networks). This library is a test collection for evaluating content-based search and retrieval of video across networks using a wide range of bandwidths. Additionally, the researchers wished to study automatic and comprehensive mechanisms for library creation. The collection contained nature, science, and news videos from a local television station and CNN. All videos were automatically segmented into short segments based on content. The collection was intended to support three 'classes' of network users. Low bandwidth users (3.8, 11.5 and 15.3 Kb s^{-1}), such as K-12 schools, had the highest M-JPEG compression ratios (150:1 and 50:1) coupled with the lowest image capture sizes (160×120 and 320×240). 'Intermediate' compression ratios of 16:1 and 50:1 along with capture sizes 320×240 and 160×120 were suggested for general audiences with (36, 46, and 144 Kb s^{-1}) bandwidth. The highest quality video was intended for users of their ATM test collection (2304, 576, and 184 Kb s^{-1}) with image capture sizes of 640×480 and compression ratios of 4:1, 16:1 and 50:1.

The Carnegie Mellon Infomedia Project built a terabyte video digital library [20,21]. This collection primarily contains news and documentaries obtained from various sources including CNN, Discovery Channel, public television productions, and public domain videos from government sources. The videos are mostly in MPEG-1, 352×240 image size, 30 fps format and are segmented with an average clip length of 1.8 minutes. For every clip, several video abstractions are available, including the video title, those that use extracted keyframes, a 'skim', and 'match bars' [7]. Researchers on this project feel that the 'wide ranging nature of the video contained in the collection has lead us to develop generic video abstractions rather than domain-specific ones applicable only to certain kinds of video' [22]. The video collection is used in sub-parts for specific research projects. Christel *et al*. [22] used 45 hours of video from three Public Television Series: The Infinite Voyage, Planet Earth, and Space Age to explore several types of video surrogates used to browse video. In this particular study there were 1600 clips, each an average of 1.7 minutes long, that were manually segmented. Other studies stemming from this project center on speech recognition, image processing and natural language processing for temporal segmentation and to extract meaningful content that can be used when indexing and retrieving videos [23,24].

Yeung and Yeo [1,25, http://www.ee.princeton.edu/~mingy/ICPR96] based their system on two types of browsing: microscopic and macroscopic. Microscopic browsing refers to browsing at the 'syntactic unit' of shots and frames. Macroscopic browsing refers to 'semantic unit' browsing of video stories and scenes. Microscopic browsing tools consist of a keyframe view of each shot and playback mechanisms. To satisfy user needs for macroscopic browsing, a method for clustering video shots into story units was implemented. The shots extracted from a video program were used in a 'scene transition graph' to represent a high-level structure of the content. The clips used to demonstrate their system were from IBM television commercials. In a previous study (1996), an episode from

the popular television sitcom ‘Friends’ was used to test and present the results of story unit clustering.

The Baltimore Learning Community (BLC) Project (http://www.learn.umd.edu) makes use of a dynamic query interface [26] for exploring resources in their digital video collection. A barfield display provides an overview of the videos in the collection. The x-axis and y-axis are categorical and can be dynamically changed to map videos according to topics, teaching standards, source (where video was obtained), or title. Keyframe surrogates, available in either ‘slide show’ or ‘storyboard’ format, provide a preview for each segment of video. The overview level allows ‘across-document browsing’ while the preview level reveals the gist of an individual segment through ‘within-document browsing’ [27]. The collection consists of 30 hours of Discovery Channel documentaries together with public domain materials from the National Archives and NASA. The videos are divided into anywhere between one and five minute length manually segmented units that are designed to be included as part of classroom lessons. Each segment designates a single distinct section of content. Furthermore, each segment is manually indexed as part of a resource collection and includes a brief textual description. All of the videos are MPEG-1 format and future plans include full-screen video streaming directly to classrooms [28,29].

Chang *et al*. [30, http://disney.ctr.columbia.edu/WebSEEk/] developed WebSeek, a searching system for images and video on the World Wide Web (WWW). WebSeek stores metadata, visual summaries and pointers to visual information sources on the WWW. The system uses Web spiders to collect information about video on the Web by detecting file extensions (such as .qt, .mpg, .avi). At the time of their publication, the system had indexed 10 000 video sequences. Indexing is accomplished by using the key terms of html tags and URLs to map the videos onto one or more subject classes in the system’s semantic ontology. WebSeek has processed more than 1 600 000 query and browse operations. Query operations that are supported are text-based, subject-based, and ‘query by example’ content-based (colour distributions). Browsing operations supported are subject navigation and visual viewing of contents.

This brief review of some current video collections shows that researchers are creating a variety of solutions to answer a wide range of video retrieval questions. They have spent a great deal of time gathering and digitizing video in order to accomplish their tasks. The types of features selected (such as amount of motion or specific genre) for the video data to be used as well as the number of hours needed vary from project to project. Collections gathered range in size from a few 1–3 minute segments to 1000 hours worth of video. Two or three 5 minute-long segments may be adequate when testing methods for automatically generating salient stills, but investigating aspects of browsing video requires a substantial set of data to work with. The collections produced have been excellent for their intended purposes and are the motivation for suggesting that collaboration to produce a test collection of video would be a useful endeavour.

1.3 *Rationale for building a test collection*

The objective of the Open Video test collection is to provide a means for researchers to begin to understand information retrieval in video collections (see http://openvideo.dsi.internet2.edu/). Previous test collections for text retrieval were built to understand the performance of retrieval algorithms and systems [13,31,32]. In these systems, recall and precision were the primary measures of performance. This test collection will be used to explore the use of precision, recall and new metrics for video retrieval. Indexing techniques for video should be explored as well as other aspects relevant to video specifically, such as segmentation into 'intellectual chunks'. The test collection will be useful for studies about searching by themes, actions and events, in addition to searching on a syntactic level by looking for features within the video such as 'brown desert with blue sky'. Within these two broad levels, researchers may explore text-based search, interactive browsing, subject search, navigation by visual summarization, or searching by example. Researchers addressing problems such as algorithms for segmenting video, various ways to browse video using surrogates and ways to automatically index video should also be able to take advantage of the Open Video testbed.

Numerous collections have already been constructed for investigating video retrieval questions. There are three motives for building yet another test collection and all underscore the benefits of creating a shared collection. The first centres on the legal aspects of building a collection, the second on community problem space and the third focuses on properties of collections. The 'legal aspects' motive is based on being able to share and distribute the test collection contents. 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 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 to replicate or make comparisons in video retrieval research; since it would not be easy to share materials used for testing with others. This is somewhat related to the 'problem space' motive for creating the test collection which is to organize a collaborative video retrieval research community. Not only will sharing results and data be less frustrating but also cooperating on the test collection construction may force researchers to form a strategy for future work in this field. The 'properties' motive for creating the test collection is to collectively work towards an effective design of the contents that avoids some of the problems seen with other test collections. For example, some 'collections' used for testing are so small that they could not be considered a representative sample. Also, it is often difficult to find explanations of the collection used for an experiment and the documentation for this test collection will be openly available.

1.4 *Overview for the test collection*

The Open Video test collection is intended to be an initial phase for a larger endeavor. This project is meant to grow as participation from the research community increases. For this first phase, we aim to assemble 500 hours of open source video data and the preliminary framework presented in this paper uses 500 hours to illustrate the mix of factors. Other video materials not in the public domain can be obtained and added later to design a collection for research purposes, much in the same way that TREC has been formed for text retrieval [33] Later, permissions may be obtained to include copyrighted materials. The framework outlined in this paper includes a set of conditions for compression, segmenting and storage issues. Additionally, an effort to explore indexing of video, definition of a set of queries along with relevance judgements, and continued work on metrics can be established during the first phase of development. With Phase I in place, researchers may begin work on indexing, searching, and browsing issues.

Although frameworks for the 'ideal' test collection have been written for text experimental collections, it is only recently that researchers have had the capability to realistically concentrate on video and other multimedia as an extension from this work. According to Sparck Jones (1976), the ideal text collection should be large and reflect real retrieval environments. Sparck Jones further states that document collections of real retrieval environments exhibit both variety and homogeneity in content, type, source, origin, range over time, and language. The documents in text experimental collections are represented by abstracts, titles, keywords, citations, author, bibliographic elements, and controlled language indexing. Although video is different from written language, some requirements for a test collection may have analogous correlates.

2. Content

The video documents that are included in the test collection should ultimately be selected according to guidelines; new documents should be added based on a plan rather than opportunistically. Without such direction, the collection may turn out to be either too narrow or too broad to be useful. The strategy for building a large collection is to begin with an initial set of specifications and start with the resources that are easiest to obtain and distribute. Public domain video will satisfy these preliminary requirements. The types of issues to consider when assembling the videos are the amount of unedited versus edited footage, content characteristics such as amount of motion, whether the videos are full length or in segments, and what compression formats are used. The test collection will be a resource for the general video IR community, thus the overall goal is to assemble video that is freely accessible in order to promote the creation, study and evaluation of video storage and retrieval systems.

2.1 *Factors*

Videos will be selected in a manner so that various factors such as genre, colour/black-white, sound/silent, and range over time will be appropriately represented in the collection. In addition to the videos themselves, several types of video surrogates will be included for phase one. There are different ways to summarize video documents for retrieval applications, including visual, audio or text elements either by themselves or in combination. The collection should ultimately contain any available text transcript files, closed captioning text files, titles, text descriptions, keyframes, as well as other available visual surrogates. The videos chosen as part of the collection should be useful in realistic situations. Some user groups will require complete video programs while others need smaller segments of programs. From a storage and distribution perspective, it is important to note that a test collection must offer the capability to transfer complete video files for use on local systems as well as the more commercially appealing streaming media solutions. Although storage requirements will increase, it is necessary to obtain an assortment of video compression formats to support multiple classes of potential users.

The challenge we face is to provide a suitable set of specifications that describe the contents of the test collection. There are many factors that need to be considered in order to maintain a test collection of video that will be both varied and specific enough to be practical. This list of factors includes: (1) genre, (2) time (both period and run length), (3) amount of motion, (4) colour or black/white, (5) sound or silent, (6) language, (7) raw footage or edited, (8) segmentation technique, (9) duration and (10) compression type. The test collection should be sufficiently large to provide videos that satisfy various combinations of these factors. For example, we do not wish to design a test collection consisting entirely of MPEG-1, black/white news clips from the 1950s. Conversely, it is not desirable to build a collection that is too scattered. Not all combinations of factor values will be possible to represent and some may not be relevant to researchers at all.

2.1.1 *Genre*. The word *genre* can be defined as ‘a type of film’ and is a category used to classify video according to certain archetypical patterns. There are numerous genres for video. Most people are familiar with the classifications used by their local video store: science fiction, drama, action, westerns, comedy, documentaries, public service, cult classics, war films and children’s videos. These mainly refer to films, but can also be applied to television. Of course, television has a further set, including game shows, talk shows, news, commercials, sports, sitcoms, public service announcements and music videos. Some of these can be broken into finer detail. For example, some of the video in ‘action’ could be split into gangster, detective and Kung-Fu stories. Another possible method would be to break the genre further according to artistic style or even by the studio that produced it. It is known that certain Hollywood studios were distinguished in the past for a particular style of filmmaking and that filmmaking techniques for the

various genres change over time. Apart from the 'professionally-directed' video, other genres exist such as training films (e.g. Military), surveillance footage, teleconference meetings, and home videos. In addition to these 'filmed' videos, there are cartoons, claymation, and computer animated programs.

The initial test collection will consist mainly of public domain video. It is expected that in the first phase of development, the collection will be limited to whatever genres can be obtained primarily from government agencies, organizations, and individuals who do not seek commercial gain from their intellectual property. It is difficult to find 'story-type' video without use restrictions. For the most part, raw uncut footage, documentaries, public service announcements, government produced news clips, historical information, training films and educational documentaries will be the most represented genres of video. As this project progresses, it will be necessary to obtain permissions so that feature films, television programs from major networks, and more recent news broadcasts may become available. The major benefit of the public domain collection is that the contents will be freely available for research purposes. However, it will be necessary to expand the collection to various genres that cannot be found for free, and these portions will be restricted to research use.

2.1.2 *Time attributes*. There are two types of time attributes. The first concerns the time period when the film was produced and the second is the duration of the video. It will be important to include video from a range of time periods to provide robust tests of algorithms across different styles and production qualities. Duration is the length of each 'unit' of video in terms of hours, minutes and seconds. Inclusion of this factor is an attempt to plan for a variety of video lengths to be added to the test collection. It will be necessary to think ahead so that many smaller length videos do not fill in all the hours allotted by the guidelines for each of the other factors.

Cinema and television history can be broken down into specific artistic or stylistic periods. Relevant to specific aspects of video retrieval research, many of these film periods were brought about by technical advances. An example would be the 'French New Wave' that started in the early 60s and was characterized by 'the creation of a new vocabulary of hand-held camera movements' [34]. This movement of cinema was shaped by the invention of a 35-mm lightweight camera in the late 50s that allowed more freedom and fluidity of movement. It would be beneficial to note the artistic period that a film belongs to, however, trying to group the videos in a test collection strictly according to these periods could get tricky. Classifying each video would require a great deal of knowledge about the artistic aspects of cinema since some of these periods overlap and many videos would not fit neatly into place. Additionally, the initial set of video may not contain very many of the types of video that would fit into artistic period categories. Therefore, it would be practical to find more general time periods for specifying the number of hours of video within each interval. Later,

as copyrighted materials are added to the test collection, it will be possible to group them according to other defined time intervals based on artistic or technical history.

The duration of each video in the collection could be anywhere from less than a minute to several hours. For each video, the duration is counted as one show, a film, or an entire portion of video that was edited at the time of production as a single program. In cases where raw footage cut from a final product is available, it will be included as pat of the final product object. If it is unedited, then the duration for the video is one continuous theme or topic of video. Each of these videos could be broken into segments with each individual segment or a subset of the segments included in the test collection.

2.1.3 *Motion attributes*. When selecting video for the collection, the amount of motion or action depicted in the video is a necessary factor to take into account. In Wolf's [6] research on keyframe selection by motion analysis, he specifically used Hitchcock's film *The Rope* because of camera techniques employed for using motion to 'direct the viewer's attention in a dramatic scene and to indicate transitions between dramatic scenes'. Most video segmentation and keyframe algorithms can produce disparate results based on the amount of motion contained in a video. It would be unwise to build an entire collection without considering whether the contents contain high amounts of motion or hardly any at all. Even though the majority of the first phase materials will be obtained from government holdings, an effort should be made to balance out the collection in terms of the amount of motion. This could be accomplished automatically through the use of motion detection algorithms to place videos into predefined categories.

2.1.4 *Colour, sound, and editing attributes*. The proportion of video in the test collection that is (1) colour/black-white, (2) silent/without sound/without dialogue and (3) raw footage/edited might seem like a negligible detail. However, these additional factors are important to many video information retrieval researchers and should not be ignored.

The techniques used for producing colour in film and television have changed a great deal over the years. Black and white films were predominately used until the mid-50s and have two dimensions: contrast and tone. Contrast is the relative darkness and lightness of the various areas of the image. Tone is the relationship between the darks and the lights. The better the stock of film used, the subtler the tone will be. Colour, was always of interest to early filmmakers. Some early black-and-white films used tinted films to give the illusion of colour. In 1935, the 'Technicolour three step process' used separate strips of film to record magenta, cyan and yellow spectrums that were then processed to produce a single colour version. The 'tri-pack' system replaced the three step with all three spectrums layered together. It was in 1952 that the first colour negative was manufactured by the Eastman Kodak Company. The use of colour has been improving ever since.

For retrieval of video, the differences in quality can affect automated processing. There are compression/decompression algorithms (codecs) that work differently with black-white videos and some are better for specific colours.

The technical capabilities for producing sound have been improving since the very first sound film in 1926. Until the development of magnetic tape in the late 40s, sound was recorded on bulky equipment that was difficult to record 'on location'. Nowadays, digital sound recordings are possible. The quality and type of recorded sound will have an influence on retrieval research through both the restrictions it imposed by limiting movement in early film to the quality of sound available for speech recognition.

It would be reasonable to own video with varying sound qualities in addition to video without sound. Video stories can be told differently, with various cinematic devices, depending on whether there is dialogue, just music or completely silent. Consider the movie 'The Red Balloon' where the entire story is conveyed without dialogue. In studies that explore methods for creating abstracts of videos, for example, researchers may find it useful to explore stories without dialogue. Omitting certain types of videos, like old silent films or other 'fragments', would deny the video retrieval community a wide variety of useful materials.

Among other uses, raw footage is necessary for investigations where it is essential to edit several versions of a film. There is evidence that recall and comprehension of video depends on both the underlying story structure as well as how its shots are sequenced [35]. By including unedited sequences, the test collection becomes more valuable by allowing investigators the opportunity to work on editing and indexing of video in ways that may help with retrieval of it later.

2.1.5 *Segmentation*. One of the factors that will determine the nature of the test collection is the amount and types of video segmenting used. The question at hand is how much of the collection should consist of full-length programs and how much should be segmented clips. Of the segmented video, what segmenting algorithms need to be represented? Moreover, should manually segmented videos along with a description of how the segments were selected be included? The advantage of using full-length video programs is that it allows researchers to control for themselves how the video is segmented. One of the drawbacks of using full-length video programs is that it is impractical to store several compressed versions of the same program. It would be more efficient to break up programs into segments and compress these clips using different compression methods. Of course, all of this may seem irrelevant if it was decided to store uncompressed video, however, this seems impractical at this time.

2.1.6 *Format*. At an image size of 640×480 pixels in 24-bit colour, one second of uncompressed video at 30 frames per second would take up about 23 Megabytes or 90 Gigabytes per hour. Needless to say, a 500 hour test collection

would require considerable disk space if stored this way. Compression is required when dealing with video documents. There are numerous ways to compress video and the techniques used will have a direct effect on retrieval systems. These effects include what algorithms are used for segmentation, how keyframes are extracted and how the user views the video. In addition to selecting a compression method, one must select parameters such as resolution, audio bit rate, and compression ratio. All of these various alternatives means that it will be useful to provide guidelines in order to prevent the situation where the entire test collection consists of videos of one compression method with the same parameters. This is important since the compression methods/parameters must be chosen carefully depending on how the video will be used. For example, if the class of intended users have Macintosh computers and 28.8 Kb s^{-1} bandwidth then it is reasonable to use MPEG-1 at 160×120 image capture size and a compression ratio of 16:1. In any case, compression methods can get outdated rapidly, updates need to be planned for and added regularly.

The type of codec (Compressor/DECompressor) chosen influences the quality of the video for the user by determining such attributes as colour depth, keyframe frequency, and compression ratio. Some codecs were created specifically for certain types of applications. For example, MPEG-4 is a low-bandwidth format suitable for Internet video and videoconferencing. There are many to select from, however, the focus for the test collection guidelines will be on the more 'popular' codecs. Sorenson, Indeo 4.4, Cinepak, and Apple codecs all fall under the QuickTime architecture and are extremely popular with commercial CD-ROMs as well as on the Internet. These QuickTime formats generally provide video of moderate quality, meaning that they are approximately low VHS standard at 15 fps with mono sound in a quarter screen window. Cinepak, for instance, was designed for 2X CD-ROMs and is typically 320×240 resolution at 15 frames per second (fps) and 180 Kb s^{-1}. Microsoft developed their own proprietary standard, Video for Windows (VFW), which also provides video of moderate quality. RealVideo, an architecture for low bandwidth users, was designed for video streaming on the Internet. RealVideo 1.0 provides two codecs: RealVideo Standard and RealVideo Fractal. Both codecs are for general purpose Internet use, but RealVideo Fractal is recommended for high bandwidth and frame rate applications such as corporate Intranets. RealVideo's codecs may be used for bandwidths that scale from 10 Kbps to over 500 Kbps and are 'tuned' for 28.8, 56 Kbps modems and ISDN. MPEG (Moving Pictures Experts Group) is an internationally agreed upon format. MPEG-1 was designed to deliver full screen video from a 2X CD-ROM and is roughly equivalent to VHS quality. Typically, MPEG-1 videos are 30 fps and are between the data rates of 140–170 Kb s^{-1}. MPEG-2 was designed for broadcast quality digital satellite, cable, high definition television as well as DVD. MPEG-2 data rates are usually 3–8 Mbit s^{-1} for Main profile (regular TV) and 15–20 Mbit s^{-1} for HDTV. MPEG-4 is a low bandwidth format for use across the Internet. It would be impossible to say which of these

is the ‘best’ format to choose. The disadvantages and advantages of each depend on their intended use within a research project.

The majority of video storage and retrieval work from research institutions in the past ten years involves the use of MPEG compressed video. Specifically, MPEG-1 is the most prevalent and has been used in cataloging, indexing and browsing research. This trend continues, although now work with MPEG-2 and MPEG-4 can be found in greater numbers. Although a complete survey of projects has not been undertaken, it is a fair guess that Motion JPEG has also been used frequently. Those studying video processing use uncompressed video. The reasoning behind the ‘MPEG’ decision for format varies. Kobla [36] sums up one explanation when he states that ‘the MPEG standard is arguably the most widely accepted international video compression standard’. Another contributing factor may be that the MPEG was developed by an organization for standards rather than by a corporation. Furthermore, MPEG provides very good quality video and was the only one around the mid-90s that provided full screen video. Nowadays, work that began with MPEG is being built-on and influential projects are using MPEG based on the numerous techniques available. The test collection content must provide the specific formats that the research community needs. Therefore, MPEG will dominate the test collection, however, other formats will be added as the demand increases.

2.1.7 *Language*. Multilingual retrieval of text, audio and video documents is an important digital library research topic [37]. Current users of electronic resources are limited to documents in languages that they are familiar with since reliable cross-language retrieval is not yet a reality. Many researchers are actively investigating methods for using the audio components of video for browsing and retrieval. Considerable work on speech recognition using English and non-English language news-broadcast videos has been completed by the Informedia Project [38,39]. Still, there is much work to be completed in this area. Inclusion of videos in the test collection that are in languages other than English will promote additional research in this area. The test collection should have a subgroup of non-English videos, and the questions that remain are: ‘Which languages should be selected?’ and ‘How much of each language type will be needed?’.

2.2 *Video composites, surrogates, and metadata*

Investigations for searching, browsing, navigating and organizing video within systems do not depend on the video data alone. Other types of information are extremely useful to own. These include the textual, audio and visual metadata about a particular video. Summarizations of video such as keyframes, salient still images, audio segments, and textual overviews would be helpful to a number of researchers. Some investigators may be excited about distributing the output from techniques they have developed, such as a new keyframe extraction algorithm.

Making these available in the test collection would be an opportunity for the output materials to be implemented in other systems and evaluated. Scripts, close captioning, annotations, and audio descriptions that are obtainable would make excellent resources.

Providing a comprehensive set of metadata that describes these video segments, surrogates, scripts, and other related resources in the repository is useful for two primary reasons. First, metadata can be used as search criteria, enabling researchers to find video segments and surrogates most appropriate for their research purpose. As the size of the video repository grows, the efficiency and effectiveness with which one can search the repository will be very important. The ability to quickly locate all video segments produced in the last five years that are in colour, include sound, and contain approximately 5000 frames, for example, might be very useful for a particular research purpose. Including a comprehensive selection of metadata for the content of the test collection, then, provides a basis for a robust search and retrieval facility.

The other primary purpose of the repository metadata is to provide researchers with as full a description of the video they are working with as possible. Although an investigator might be primarily interested in video segments that are in colour and of approximately 5000 frames in length and use those attributes as search criteria, after the investigator begins working with the video, other attributes, such as the data rate of the video or how it was segmented, may be important. Providing this data as part of the repository means it only has to be determined and recorded once, and reduces the burden on researchers to extract relevant metadata detail from the video content themselves.

A relatively small set of metadata describing the video content currently contained in the Open Video Project repository has been collected and entered into a database. Tables 3, 4, and 5, shown in the Implementation section of this paper, list the metadata currently included the database. This metadata is of three types: attributes of the source video (for example, when it was produced, who produced it, what sort of content it contains), attributes of the segment digitized from the source video (duration, number of frames, compression method), and attributes of available video surrogates and composites (the type of surrogate, its duration, how its frames were extracted). Although the attributes listed in the tables are the ones we are currently using in the database, we do not consider this to be the definitive set of metadata for the test collection; we are seeking suggestions from potential users of the repository as to what metadata would be most important to include in the database and will make adjustments as appropriate.

While the database is in the beginning stages of development, we believe that the speed of development will increase if researchers are able to actually start using the test collection. To this end, a publicly available Web-based interface to the database (available from http://openvideo.dsi.internet2.edu) has been created. This interface enables a researcher to search the contents of the collection by several attributes. The results of the search are presented in a summary form,

with access to both complete details about the segment (all metadata) and the segment itself one mouse click away. Both the search and results display of this interface will be refined as suggestions are gathered from the digital video research community.

2.3 *Metrics*

Text-based document collections typically use two measurements, recall and precision, that are designed to judge how well users are able to obtain information from a particular retrieval method. Recall is the proportion of relevant documents that are actually retrieved and is defined as the number of relevant items retrieved divided by the total number of relevant items in the collection. Precision is the proportion of retrieved documents that are actually relevant and is defined as the number of relevant items retrieved divided by the total number of items retrieved. For the text test collections, the aim is to determine which technique or combination of techniques results in the best precision/recall figures. Recall and precision can also be used as measurements for video collections. However, the concepts of document (segment? entire video?) and corpus (final products only? metadata? out takes?) must be specified. A test collection for video will be a basis for discussion about precision and recall as well as an opportunity to better understand these measurements for video collections.

Besides precision and recall, video collections may generate other types of metrics. Some of these have been explored in studies of browsing video with keyframes. 'Object recognition' measures user's ability to browse a set of keyframes with only a fuzzy idea of what they need [9]. This was considered different from 'object identification' that gauges the user's ability to search for a specific object. 'Gist determination' estimates a user's ability to understand the theme of a video document by browsing a set of keyframes [11,40,41]. An 'action recognition' measurement [41] judges whether people are able to determine the video's action content or 'identify multiple objects' from a set of keyframes well enough to select the relevant video. These types of metrics will play a role in future video IR research and they will eventually come into consideration by researchers that use the test collection.

Retrieval of video documents from a collection can be achieved in a number of different ways. Videos can be searched using (1) text, either in natural language or with keywords, (2) visual feature (still images or motion video) and/or (3) by auditory feature queries. A goal of the test collection will be to study the characteristics and behaviour of various query formulations. Queries might be expressed in any of the formats listed above or by combining them. On the other side, results sets can be displayed using any of the input formats or combinations. Building a test collection of this size will provide a useful set of videos for researchers who are working on interfaces for visualizing and displaying result sets.

3. Implementation plan

Using the 500 hour example, one minute of compressed video requires about 10 MB of hard disk space, so 500 hours will claim 300 000 MB or about 293 GB. Additional space is needed for video that is stored in multiple formats or segmentation solutions. Further space will be needed to hold video metadata. It is safe to assume a terabyte storage problem for such a test collection. To begin exploring the feasibility of such a testbed, the Open Video Project has assembled some video from US government (including some contributed by the Informedia Project) and non-profit organization sources and begun experimenting with ways to organize and deliver it. At the time of writing, about 20 hours of MPEG-1 video in several genres are available through a special channel provided by the Internet2 Distributed Storage Infrastructure Initiative (http://dsi.internet2.edu) that supports distributed hosting for research and education in the Internet2 community (see other papers in this special issue). At present, the primary data files are stored on the distributed I2-DSI backbone and the metadata (in an Access database) and retrieval interface are available on a server at the University of North Carolina. From the retrieval interface, users can download selected video files from the I2-DSI servers in an efficient manner. That is, the I2-DSI system provides dynamic redirection of URL requests so that the download takes place from the closest I2-DSI replication server relative to the client.

Preliminary provisions for adding contributions from the research community (either the metadata alone or both primary and metadata files) are in place so that others may place video in the repository or provide links to video data they maintain themselves. It is hoped that by providing flexible options for other institutions to participate, the research community will grow quickly and effectively, as have other open source communities (see [42] for a study of the Linux community). The recommendations below are meant to be provocative and illustrative rather than definitive. Clearly, a global research community will require better proportions of language coverage and new technical advances may require additional segmentation and coding schemes. For the present, these recommendations offer a starting point for discussion.

3.1 *Recommendations*

A preliminary list of recommendations for the amounts of each type of video is supplied in Table 1. These suggestions are a first guess at what the test collection should contain and are intended to stimulate discussion. After responses are received, the list will be modified to reflect what is most needed to carry out video retrieval research. In order to better understand the decisions made when creating this list, a short summary of the rationale behind each choice will be provided for each factor.

Table 1. *Preliminary composition estimates for the open video repository*

Factor	Types	Amount (%)	Amount (hours)
Genre	Documentary	20	100
	News	20	100
	Fictional video, comedy	2	10
	Fictional video, drama	2	10
	Fictional video, action	2	10
	Public service	2	10
	Propaganda	8	40
	Historical record keeping	8	40
	'Home' movies	8	40
	Surveillance	8	40
	Sports	10	50
Time period	1896–1912	4	20
	1913–1927	8	40
	1928–1932	15	75
	1933–1946	15	75
	1947–1960	20	100
	1961–1970	12	60
	1971–1980	12	60
	1981–1990	8	40
	1991	6	30
Amount of motion	Low	30	150
	Medium	40	200
	High	30	150
Colour/B-W	Colour	85	425
	Black/white	15	75
Sound	With music only	5	25
	With dialogue only	5	25
	Sound	75	375
	Silent	15	75
Language	English	94	470
	Spanish	2	10
	German	2	10
	French	2	10
Raw/edited	Unedited footage	20	100
	Edited	80	400
Duration	Over 4 hours	10	50
	3–4 hours	8	40
	2–3 hours	8	40
	1–2 hours	15	75
	30 minutes–1 hours	15	75
	15–30 minutes	20	100
	5–15 minutes	18	90
	Up to 5 minutes	6	30
Segmentation	No segmenting	30	150
	Manual	30	150
	Based on scene change	25	125
	Based on fixed time length	15	75
Compression	MPEG-1	30	150
	MPEG-2	30	150
	MPEG-4	20	100
	M-JPEG	10	50
	Other	10	50

3.1.1 *Genre*. The genres listed for the first 500 hours cover only a small subset of all available categories for video content. Definitions for each of the genres included in Phase 1 are listed in Table 2. After reviewing the selection of video that is in the public domain, the proportions were based largely on what we anticipate will be available. To illustrate, only about 6% of the 500 hours will consist of 'fictional video' since there is not much of it that is obtainable.

3.1.2 *Time period.* The time periods were chosen based roughly on 'technical' and artistic film history until the 60s. From 1896–1912 marks 'the evolution of film from sideshow gimmick into economic art' [34]. The era of silent films spans from 1913–1927. There are a number of these early films, most no longer have use restrictions. Technologically significant improvements in terms of sound and quality, in addition to stylistic changes occurred during 1928–1932. The 'great age of Hollywood' film fell between 1933–1946, and is also a period of advances due to World War II. From 1947–1960, the film industry confronted the challenge of television and there was a growing internationalism of film art. After the 60s, we separate time periods by decades; 60s, 70s, 80s, and 90s to the present. Thirty-eight percent of the hours allocated for the test collection consist of video that was created after the 60s, which may be challenging to obtain due to copyright limitations. For video produced after 1990, it becomes even more difficult to obtain interesting and useful edited public domain materials. We might find that a large portion of this 6% is raw unedited footage.

Table 2. *Genre definitions list*

Documentary: A presentation of factual information intended for entertainment and/or educational purposes.
News: Any edited or unedited factual information that includes a commentator providing reports about recent events (at the time of video creation).
Fictional video: Any fictional work or a story that is conveyed to the audience. These generally, although not necessarily, contain elements such as a setting, characters and a plot. This super-category contains television shows such as sitcoms and feature films. Since it is not expected that much of this type will be available in the public domain, they have been grouped together. We have selected three general types: comedies, action and drama. Action films include any war films that are fictional in nature. As an example, D.W. Griffith's classic films 'the Birth of a Nation' would fall under 'fictional video, drama'.
Public service: Any announcements about safety, government programs or benefits that would serve the population.
Propaganda: Video that was created in order to spread ideas and information to further or damage a cause.
Historical record keeping: Video that was created with the intention of recording some event in history. Includes presidential speeches, demonstrations and natural disasters. This genre is different from 'news' since it was not meant as a 'report' on the topic.
Home movies: Amateur, mostly unedited clips that depict ordinary life.
Surveillance: Uncut and unedited footage taken from cameras that are set up to observe and monitor an area.
Sports: A video (with or without commentary, raw or edited) that depicts physical exercise and/or team games.

3.1.3 *Amount of motion*. The ‘amount of motion’ factor for video is split into three categories: low, medium and high. Low motion means that there is hardly any action, perhaps a person sitting at a desk talking. A ‘medium motion’ video is one that contains some inactive intervals as well as some periods of high action. It can also mean that the video has even amounts neither low nor high action throughout. High motion videos are those that contain many action scenes with a great deal of frequent movements, such as a football game, a person running or a bird in flight. A range of values for each of these categories should be established for use with automated motion detection methods. Automation can be a more efficient procedure for determining whether a particular segment of video has low, medium or high motion.

3.1.4 *Colour/black-white*. Colour videos are all those that were processed either with colour film or dye-transfer printing. This category includes Technicolour, the ‘tri-pack’ system, and colour negatives. A number of black-white films made in the 20s used tinted stock to provide a dimension of colour. These are different than ‘colour’ and if we acquire them, they should be considered as ‘black-white’. Not too many videos are part black-white and part colour, so we will not include these. If there are enough of them, they can be added to Phase II. Black-white videos are those that were intentionally created without colour.

3.1.5 *Sound*. Sound was not usually recorded along with film until the late 20s although it was technically possible in 1919. Thus, the 15% allotted for silent film covers the 12% for the first two time periods and an additional estimated 3% from other time periods. These could come from home videos or other raw footage. Silent film was created without music or dialogue. Although many silent films had musical scores that accompanied them, unless the score has been dubbed with the film on video, these are considered as ‘silent’. Any available music score, either digitized audio or as an image file will be included as metadata. The ‘with music’ sound category includes video that was created with music but no human dialogue. The ‘with dialogue’ category is just the opposite, containing speech without any music. All videos that contain both speech and music are considered ‘sound’ videos. It is anticipated that the largest proportion of video will contain both music and dialogue.

3.1.6 *Language*. In Phase I of this project, English language (non-dubbed) videos will be the most prevalent. We acknowledge that as the collection grows, a larger sized non-English sub-collection should be assembled. The languages suggested were chosen based several searches of databases for public domain video collections (such as the National Archives NAIL database). A small percentage of Spanish, German and French (also non-dubbed) videos will be included for cross-language research. These languages are those that are most widely used and are most easily obtainable. The addition of Asian languages was

explored (Japanese, Korean and Chinese). These should be incorporated in the next Phase of the test collection development since it was more difficult to locate videos in these languages.

3.1.7 *Raw/edited*. An edited film is one that has been assembled by purposely cutting and rearranging individual camera shots into larger meaningful units. Historical record keeping and home videos are expected to be entirely raw unedited footage. Twenty percent of video is allocated to unedited clips, 18% from these two genres with an extra 2% from other genres.

3.1.8 *Duration*. Duration refers to a video's length before it has been segmented. Hour/minute intervals form eight separate categories. We tried to make a 'best guess' when specifying the proportion of video allotted to each category. Specifically, surveillance video is usually very long and will take up the majority of the 'over 4 hour' category. Also, there are numerous shorter length public domain clips and these will probably account for the 30% non-segmented. The 'duration' guidelines will need to be flexible due to the fact that it will be almost impossible to result in the precise number of hours proposed.

3.1.9 *Segmentation*. We chose four broad categories for video segmentation. An un-segmented video is an entire, intact file consisting of one unedited topic or edited program. Manually segmented video has been cut, based on human judgment, into meaningful content-based units. Video segments 'based on scene changes' are those that were segmented using any technique that breaks up video into chunks based on automatically detected scene changes. All algorithms for scene change detection are grouped into this one category. Phase II will include more specific guidelines, breaking this category into smaller ones. Videos segmented 'by time length' are either manually or automatically segmented at regular time intervals, for instance every 3 minutes or every 5 minutes.

3.1.10 *Compression*. MPEG compressed video dominates in this collection due to research demands. Eighty percent of video in the collection will be either MPEG-1, MPEG-2, or MPEG-4. Ten percent will be compressed using Motion JPEG, another popular standard. Another 10% will consist of any other formats, such as QuickTime or RealVideo. Within each of these compression formats, there are numerous combinations of parameters. For Phase I these will be undefined, however, we will break them into ranges during Phase II. Uncompressed video was excluded from Phase I due to the high amounts of space needed for storage.

3.2 *Metadata*

A metadata database will store attribute information for each video in the collection. The purpose of this database is to provide comprehensive descriptive information about each video that users of the video repository can refer to when

they are searching for appropriate video segments for their research. Information describing characteristics of the source video, the digitized video segment, and any available surrogates or supplementary materials for each video will be included in the database. Specific attributes proposed to be stored are listed in Tables 3, 4, and 5.

3.3 *Community*

Establishing a test collection may provide the necessary resources, however with collaboration from a community of researchers even more can be achieved.

Table 3. *Source video attributes*

Attribute name	Description
Unique ID	Unique identifying key, automatically generated
Primary title	Main title of video (e.g. The name of the video tape itself)
Secondary title	A secondary title, to distinguish the specific segment, if necessary
Production date	Date the original video was produced
Creating org	Source of original video
Genre	Type of content (documentary, new, fictional, public service, etc)
Description	General description of video content
Colour	Yes (video is in colour) or No (video not in colour)
Sound	Music only, Voice only, Sound, Silent
Amount of motion	Low, Medium, High
Copyright	Copyright statement, statement of use

Table 4. *Digitized video attributes*

Attribute name	Description
Digitizing org	Organization responsible for digitizing original video source
Digitization date	Date the video segment was digitized
File size	Size of digitized video, in MB
Compression format	MPEG-1, MPEG-2, MPEG-4, Motion JPEG, Real Video, Quick Time
Duration	Length of digitized video in seconds
Frame dimensions	Pixel dimension of frame in width $\times$ height
Number of frames	Total number of frames in video
Frame rate	Frames per second (fps)
Compressed data rate	Data rate in Mbps per second
Codec	Specific compression scheme used to digitize
Edited	Yes (original video edited when digitizing), No (original digitized in its entirety)
Segmentation	How video was segmented: None, Manual, Scene change, Fixed time
Location	Complete URL to video file
Associated material text	Information about surrogates, pointers to other compressed versions of the video, or other related material
Associated material file	Complete URL to associated material file
Contributor name	For contributed video segments—Name of the contribution
Contributor org	For contributed video segments—Organization of the contributor
Contributor email	For contributed video segments—E-mail address of the contributor

Table 5. *Surrogate attributes*

Attribute name	Description
Unique ID	Unique identifying key (matching Unique ID in Video table)
Number	Number of surrogate frames created
Extraction method	Method of extracting surrogate frames
Frame size	Pixel dimensions of surrogate frames

Open source communities have led to new engineering and economic models for software and it is time that similar efforts are made for information resources. In many respects, the digital library movement will ultimately move in this direction, [43]. In addition to papers and web sites to encourage participation, a repository was discussed at the SIGIR workshop on video retrieval in Berkeley in August of 1999 and a symposium was hosted in October 1999 in Chapel Hill to discuss video retrieval evaluation and a test repository (see http://ils.unc.edu/idl/events/Symposium_Overview.html for details). We will continue to encourage such discussions and promote cooperation between researchers for improving the test collection. The guidelines that have been suggested in this paper are not 'set in stone' but rather meant to be flexible and serve as the basis for more permanent specifications. In terms of supplementing the test collection, the community should undertake two essential tasks. The first concerns keeping track of and describing new metrics. The list of metrics described in the preceding Section 2.3 must be expanded, and additional research in this area is hoped for as use of the test collection grows. The second task involves building query collections, a topic not considered in this paper. To begin to study aspects of searching video, it will be necessary to provide at least one query collection that consists of user supplied queries and relevance judgements.

3.4 *Phase II*

Putting the first 500 hours of public domain video in place is only the first step to a successful test collection. Unfortunately, the available types of 'good quality' public domain video are limited. The majority of video documents are copyrighted and it will be essential to include these as well. The second phase of the project should focus on acquiring the legal authorization needed to make use of resources that have copyright restrictions. Another topic for further discussion is the addition of a subgroup of 'non-English' video. This 'subgroup' might be one aspect to focus on specifically by those involved in cross-language video retrieval. The development of query collections and metrics for gauging how well users are able to locate wanted video documents is yet another area of inquiry that should be expanded after the first phase of the test collection is fully developed.

4. Conclusion

In this paper, we have presented the motives for building a shared video test collection. One benefit to building an open source collection is that researchers will be free of the constraints of copyrighted materials as long as they use them for non-commercial purposes. This will allow unrestricted use of experimental video retrieval systems and will facilitate collaborative work. Another advantage comes through the cooperative construction of the test collection. It will force the community to form a strategy or 'problem space' for conducting research. The test collection can be used to study a wide range of problems or focus on specific aspects depending on the directions chosen. There are also gains that can be achieved on the size and other properties of the collection. Collectively, it is possible to construct a very large collection more quickly than working alone. Metadata and documentation for the database can be richer because many groups will be working with the same data. This paper overviewed an assortment of video retrieval projects in order to illustrate the variety of exciting questions and problems to be solved in the area of video retrieval. A thoughtfully designed test collection will be instrumental in helping investigators reach their goals. Additionally, the existence of a large collection of video will enable Internet2 researchers to investigate distribution, replication, and storage issues for the Internet of tomorrow.

The suggested test collection framework outlined in this collection requires discussion and feedback to refine. We suggested a number of different factors that should be considered when building a test collection. These were (1) genre, (2) time period, (3) amount of motion, (4) colour or black/white, (5) sound or silent, (6) language, (7) raw footage or edited, (8) segmentation technique, (9) duration and (10) compression type. Are these really important aspects to consider in building the test collection? Should some of these factors be modified or removed? Are there other factors that need to be added? Are the numbers of hours specified for each factor appropriate? The paper has only touched briefly on sources for public domain video. What other sources, besides government and home-made videos, are out there and does this change the initial plan? The paper has also neglected to include some very practical considerations. Who will manage and maintain the test collection, metadatabase, and query collections? We do not yet have a cost estimate for this project. It will be necessary to acquire and sustain funding. In summary, we have designed a plan for a test collection that has the potential to become a valuable tool to the video retrieval community; however, it cannot be undertaken without combined efforts.

This discussion of a test collection illustrates a number of issues of interest to researchers developing distributed storage solutions. First, facilities for transferring files as well as streaming video are required. Second, there may be substantial redundancy in large files of the same content using different compression schemes.

Third, there will be strong coupling of files with metadata—which itself may have many variations depending on the user's needs. Finally, video retrieval researchers will need to upload results of experiments that themselves include large files (e.g. the results of segmenting or indexing a set of videos).

References

1. M. M. Yeung & B. Yeo 1996. Time-constrained clustering for segmentation of video into story units. Paper presented at the *International Conference on Pattern Recognition* 1996.
2. V. Kobla, D. Doermann & C. Faloutsos 1997. Video trails: Representing and visualizing structure in video sequences. Paper presented at the *ACM Multimedia 97*, Seattle, Washington.
3. A. Dailianas, R. Allen & P. England 1995. Comparison of automatic video segmentation algorithms. Paper presented at the *SPIE-Phototonics East '95*, Philadelphia, Nov. 1995.
4. P. Aigrain, P. Joly & V. Longueville 1995. Medium knowledge-based macro-segmentation of video into sequences. Paper presented at the *ISCAI '95*.
5. R. Lienhart, S. Pfeiffer & W. Effelsberg 1997. Video abstracting. *Communications of the ACM* **40**(12), 55–62.
6. W. Wolf 1996. Key frame selection by motion analysis. *IEEE* 1228–1231.
7. M. Christel, D. Winkler & C. R. Taylor 1997. Improving access to a digital video library. Paper presented at the *Human-Computer Interaction: INTERACT97, the 6th IFIP Conference on Human-Computer Interaction*, Sydney, Australia.
8. H. Aoki, S. Shimotsuji & O. Hori 1996. A shot classification method of selecting effective key-frames for video browsing. Paper presented at the *ACM Multimedia '96*, Boston, MA, 1–10.
9. T. Tse, G. Marchionini, W. Ding, L. Slaughter & A. Komlodi 1998. Dynamic key frame presentation techniques for augmenting video browsing. Paper presented at *AVI '98*, Italy.
10. W. Ding 1999. *Cognitive Processing of Multimodal Surrogates for Video Browsing*. Unpublished doctoral dissertation, University of Maryland, College Park, MD.
11. L. Slaughter, G. Marchionini & B. Shneiderman 1997. Comprehension and object recognition capabilities for presentations of simultaneous video key frame surrogates. Paper presented at the *Research and Advanced Technology for Digital Libraries: ERCIM European DL Conference*, Pisa, Italy, 41–54.
12. K. Sparck Jones & C. J. Van Rijsbergen 1976. Information retrieval test collections. *Progress in Documentation*, **32**(1), 59–75.
13. D. Harman 1993. The DARPA TIPSTER project. *SIGIR Forum* **26**(2), 26–28.
14. E. Voorhees & D. Harman (eds.) 1998. *The Sixth Text Retrieval Conference (TREC-6)*. NIST Special Publication 500-240. Gaithersburg, MD.
15. C. Schmidt & P. Over 1999. Digital video test collection. Paper presented at the *ACM SIGIR Conference on Research and Development in Information Retrieval*, Berkeley, California.
16. T. Chau & L. Ruan 1995. A video retrieval and sequencing system. *ACM Transactions on Information Systems* **13**(4), 373–407.
17. H. J. Zhang, C. Y. Low, S. W. Smoliar & J. H. Wu 1995. Video parsing, retrieval and browsing: An integrated and content-based solution. Paper presented at the *ACM Multimedia '95*, 15–24.
18. W. Wolf, Y. Liang, M. Kozuch, H. Yu, M. Phillips, M. Weekes & A. Debruyne 1996. A digital video library on the world wide web. Paper presented at the *ACM Multimedia '96*, Boston, MA, 433–434
19. S. Gauch, W. Li & J. Gauch 1997. The VISION digital video library. *Information Processing & Management* **33**(4), 413–426.
20. H. D. Wactlar, T. Kanade, M. A. Smith & S. M. Stevens 1996. Intelligent access to digital video: Informedia project. *Computer*, **29**(5), 46–52.
21. H. D. Wactlar, M. G. Christel, Y. Gong & A. G. Hauptmann 1999. Lessons learned from building a terabyte digital video library. *Computer*, **32**(2), 66–73.

22. M. Christel, D. Winkler & C. R. Taylor 1997. Multimedia abstractions for a digital video library. Paper presented at the *ACM Digital Libraries '97 Conference*, Philadelphia, PA.
23. A. G. Hauptmann, M. J. Witbrock 1997. Informedia: news on demand—multimedia information acquisition and retrieval. In *Intelligent Multimedia Information Retrieval*. (M. Maybury ed.). Menlo Park, CA: MIT Press.
24. M. Smith & T. Kanade 1996. *Video Skimming for Quick Browsing based on Audio and Image Characterization* (CS Technical Report CMU-CS-95-186R). Pittsburgh: Carnegie Mellon University.
25. M. Yeung & B. Yeo 1996. *Video Visualization for Compact Presentation of Pictorial Content* (IBM Research Report RC 20615): IBM.
26. C. Ahlberg & B. Shneiderman 1994. Visual information seeking: Tight coupling of dynamic query filters with starfield displays. Paper presented at the *ACM SIGCHI Human Factors in Computing Systems*, New York, N.Y.
27. G. Marchionini 1995. *Information Seeking in Electronic Environments*. Cambridge, UK: Cambridge University Press.
28. A. Rose, W. Ding, G. Marchionini, J. Beale & V. Nolet 1998. Building an electronic learning community: From design to implementation. Paper presented at *ACM SIGCHI Human Factors in Computing Systems*, Los Angeles, CA, 203–210.
29. G. Marchionini, V. Nolet, H. Williams, W. Ding, J. Beale, A. Rose, A. Gordon, E. Enomoto & L. Harbinson 1997. Content+Connectivity=Community: Digital resources for a learning community. *Proceedings of ACM DL '97*. (Pittsburgh, PA, July 23–26, 1997), pp. 212–220.
30. S. Chang, J. Smith, M. Beigi & A. Benitez 1997. Visual information retrieval from large distributed online repositories. *Communications of the ACM* **40**(12), 63–71.
31. B. Masand & C. Stanfill 1993. An information retrieval test collection on the CM-5. Paper presented at *TREC '93*, 117–122.
32. E. Fox 1983. *Characteristics of Two New Experimental Collections in Computer and Information Science Containing Textual and Bibliographic Concepts* (Computer Science Department Technical Report 83-561). Cornell: Cornell University.
33. D. Harman 1995. The TREC conferences. Paper presented at the *Hypertext—Information Retrieval—Multimedia: Synergieeffekte Elektronischer Informationssysteme*, Konstanz, Germany: Universitaetsforlag Konstanz, 9–28.
34. J. Monaco 1977. *How to Read a Film: The Art, Technology, Language, History, and Theory of Film and Media*. New York: Oxford University Press.
35. P. S. Cowen 1988. Manipulating montage: Effects on film comprehension, recall, person perception, and aesthetic responses. *Empirical Studies of the Arts* **6**(2), 97–115.
36. V. Kobla, D. Doermann & A. Rosenfeld 1996. *Compressed Domain Video Segmentation* (CAR-TR-839, CS-TR-3688). College Park: University of Maryland.
37. D. Oard 1997. Serving users in many languages: Cross-language information retrieval for digital libraries. *D-Lib Magazine, December 1997*, http://www.dlib.org/dlib/december97/oard/12oard.html.
38. A. G. Hauptmann 1995. Speech recognition in the informedia digital video library: Uses and limitations. Paper presented at the *ICTAI-95 7th IEEE International Conference on Tools with AI*, Washington, DC, Nov. 6–8, 1995.
39. P. Guetner, M. Finke, P. Scheytt, A. Waibel & H. Wactlar 1998. Transcribing multilingual broadcast news using hypothesis driven lexical adaptation. Paper presented at the *DARPA Broadcast News Transcription and Understanding Workshop*, Lansdowne, VA. Feb 8–11, 1998.
40. W. Ding, G. Marchionini & T. Tse 1997. Previewing video data: Browsing key frames at high rates using a video slideshow interface. Paper presented at the *International Symposium on Research, Development, and Practice in Digital Libraries*, Tsukuba, Japan.
41. A. Komlodi & G. Marchionini 1998. Key frame preview techniques for video browsing. Paper presented at the *ACM Digital Libraries*, Pittsburgh, PA, 118–125.
42. B. J. Dempsey, D. Weiss, P. Jones & J. Greenberg 1999. *A Quantitative Profile of a Community of Open Source Linux Developers*. (SILS Technical Report TR-1999-05) Chapel Hill: University of North Carolina.

43. G. Marchionini 1999. Augmenting library services: Toward the sharium. *Proceedings of International Symposium on Digital Libraries 1999* (Tsukuba, Japan, September 28–29, 1999). 40–47.
K. Sparck Jones 1975. A performance yardstick for test collections. *Journal of Documentation*, **31**(4), 266–272.