A Digital Video Archive System of NDAP Taiwan

Hsiang-An Wang, Guey-Ching Chen, Chih-Yi Chiu, and Jan-Ming Ho

Institute of Information Science, Academia Sinica, Taipei, 115, Taiwan
{sawang, ching64, cychiu, hoho}@iis.sinica.edu.tw

Abstract. The National Digital Archives Program (NDAP), Taiwan has developed advanced technologies for managing digital video archives. The technologies enable us to build indexing systems for fast retrieval of digital video contents, and add values to the contents. This paper takes the Digital Museum of Taiwan's Social and Humanities Video Archive project as a case study to demonstrate the role of information science technologies in developing digital video archive systems and digitizing video and audio resources. By sharing our experience and the technologies developed in our research, we hope to provide digital content providers and researchers with guidelines for the design and development of digital video archive systems and value-added video/audio data.

Keywords: video archive, video content analysis, video index, video management.

1 Introduction

The National Digital Archives Program (NDAP) of Taiwan, which was launched on January 1, 2002, is sponsored by the National Science Council (NSC). The program's objective is to promote and coordinate the digitization and preservation of content at leading museums, archives, universities, research institutes, and other content holders in Taiwan [14].

The Digital Museum of Taiwan's Social and Humanities Video Archive is an applied research project of NDAP's video and audio archives. Its main purpose is to offer free public access to a digital library containing 3200 volumes (1600+ hours) of 16mm and Beta cam video footage. This video content was produced or collected by Daw-Ming Lee, at Taipei National University of the Arts (TNUA) [4] [10].

The archive is a collaborative project between TNUA and Institute of Information Science (IIS), Academic Sinica. TNUA is responsible for digitizing video/audio data, constructing metadata, user interfaces, and the visual presentation of information. Meanwhile, IIS is responsible for providing and integrating information technologies, and building metadata databases and management systems. IIS is also responsible for developing the following sub-systems: video/audio data format transformation, shot detection, metadata searching, audio searching, and the environment for the integration and distribution of data streaming.

The remainder of the paper is organized as follows. Section 2 describes related works. Section 3 describes the video archiving process. Section 4 details the system

S. Sugimoto et al. (Eds.): ICADL 2006, LNCS 4312, pp. 42 – 50, 2006.

architecture, its implementation, and addresses several implementation issues. Finally, we present our conclusions and indicate some future research directions in Section 5.

2 Related Works

Many digital archiving systems (DAS) have been developed since the mid 1990s. The goal of such systems is to provide user-friendly ways to save and present digital content so that users can retrieve and browse it easily.

Recently, there has been a rapid growth in video content produced by traditional means (e.g., news channels, educational content, entertainment media), and individuals. Consequently, many DAS have gradually extended their archived material from text/image content to video content. However, building a video archiving system is extremely challenging due to the size of the files and the content indexing problem. A number of researchers have presented various techniques for, and shared their experience in, building better video archiving systems.

The Informedia project [1] is famous for developing new technologies for video library systems. It uses a combination of speech, language, and image understanding to segment and index a linear video automatically. A speech recognizer is used to automatically transcribe a video soundtrack into text information, and a "video skimming" technique creates a video abstract that facilitates accelerated viewing of video sequences.

Another important video management project is IBM's CueVideo [5], which uses shot-boundary detection to summarize a video and extract key frames. It acquires spoken documents from videos via a speech recognition component, and the transcribed text is indexed to retrieve related audio/video clips.

In 2002, Marchionini and Geisler published the Open Video Digital Library (OVDL) [9], an integrated system that processes data for digital video archives. In this system, key frames are first extracted using MERIT software [12] and a Java program. Then, keywords are annotated, mainly manually, for the video and audio content. In addition, OVDL catalogs videos based on the attributes of genre, duration, color, and contributing organization. It also combines a number of key frames into a storyboard in order to present video content rapidly.

The Físchlár System [8] is an ongoing project that Dublin City University (Ireland) began developing in 1999. It utilizes advanced technologies for video management and analysis. First, it detects video shots via a shot-boundary detection module. Second, it deletes advertisements from the video shots obtained in the first step. In the third step, the system analyzes the content of remaining shots by spoken dialogue indexing, speech/music discrimination, face detection, anchorperson detection, shot clustering, and shot length cue, all of which are implemented based on the SVM algorithm. Finally, it applies the story-segment program to combine several shots into a story segment and saves the result in the database.

The systems and projects described above provide good guidelines for building a digital video library; however, they only process Western languages. Until recently, there has been a lack of techniques and experience for developing a video archiving

system (e.g., a speech recognizer and a caption recognizer) for a Chinese environment. In this work, we report on such a system and release source code of two components, the watermark appending module and the format transformation module. The source code and executable program are available on the Open Source Software Foundry [15]. We discuss the components further in Section 3.2.

3 Video Archiving Process

Our video archiving process is divided into two stages. In the first stage, we choose an appropriate metadata standard to preserve the detailed description of our video file. The second stage is video digitization and content analysis, in which we digitize Betacam tapes into digital files and send the digital videos to the content analysis modules. This process reformats Betacam tapes into a digital format so they can be managed by our digital video archive system (DVAS).

3.1 Metadata Analysis

A number of video metadata standards have been proposed, for example, MPEG-7, developed by the Moving Picture Experts Group; the Standard Media Exchange Framework, developed by the BBC; the P/Meta Metadata Exchange Standard, developed by the European Broadcasting Union; the European CHronicles On-line project (ECHO), developed by the European Community [6]; and the Dublin Core application profile for digital video, promoted by the Video Development Initiative [16].

In our research, we initially used the Dublin Core metadata standard as a guideline to analyze the metadata. We found that, although the basic columns fulfill the needs of content description, the 15 columns defined by Dublin Core are not sufficient to describe all the content properties required in the management and archiving of audio and video content. Thus, in the second stage of our project, we used the metadata standard developed by ECHO as our guideline for metadata analysis because its definition of video metadata is more detailed than that of Dublin Core. The ECHO standard is an adaptation of the Functional Requirements for Bibliographic Records Model (FRBR Model) of the International Federation of Library Associations and Institutes (IFLA). We made minor modifications to the ECHO metadata standard in order to analyze, design, and develop the metadata management system for our digital archives and databases.

3.2 Video Digitization and Content Analysis

In this stage, we first transfer Betacam tapes to MPEG-2 files via a video capture card so that we can analyze, process, and preserve the video content at a later stage. Fig. 1 shows the video analysis and processing procedure, which is divided into six modules, namely: metadata injection, caption recognition/appending, voice recognition, shot detection, watermark appending, and format transformation. As these modules are all independent, users can utilize multiple computers to access different modules to reduce the processing time. In addition, all of these modules support batch operation to process a large number of video files in one operation.

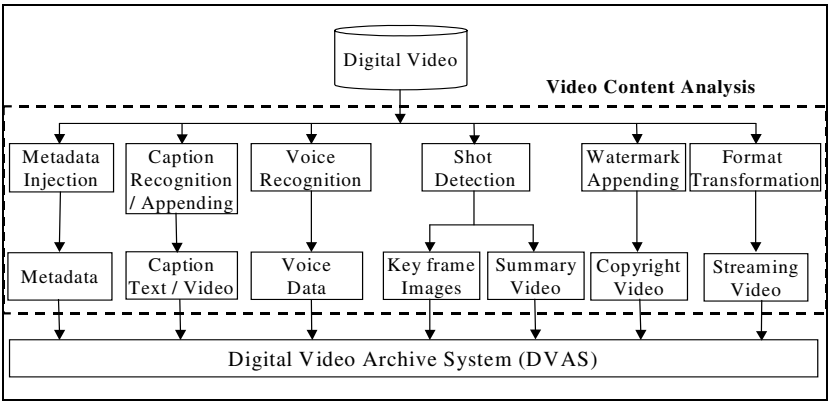

Fig. 1. The video analysis and processing procedure

1. The metadata injection system is a web system that is connected to the database system of DVAS. It allows a content provider to input metadata about a video via the user interface. The injected metadata is saved in the database and used to support DVAS when responding to users' keyword queries. To prevent misuse of the metadata by unauthorized people, the system contains a member authentication mechanism.

2. Caption recognition/appending. (i) The caption recognition module automatically retrieves the transcript from video content with captions if the content provider does not supply a transcript. Temporal information for indexing video content is also saved. A user can then use a keyword search to browse related video clips via the transcript and extra temporal information. In order to adapt to the general language of video files in Taiwan, this module focuses on processing Traditional-Chinese captions. The module was developed by joint cooperation between Chang et al. [7] and us. Its accuracy rate for recognition is over 90%. (ii) The caption appending module appends captions to a video from an external text file if users want to add captions to an uncaptioned video.

3. The voice recognition module is similar to the caption recognition module, but it processes the audio content of a video file. The module is developed via cooperation between Wang et al. [10][13] and our laboratory. It retrieves the transcript from the audio channel of the video, and saves it in DVAS. Users can use a keyword to search video content via the transcript, and then browse related video segments. This module, which focuses on processing Mandarin Chinese speech for videos, has an accuracy rate between 40% and 95%, depending on whether the voice data is noisy or clear.

4. The shot detection system performs shot detection on the MPEG1/2 files and outputs the analysis results as an XML file containing the temporal information about locations where scene content changes dramatically. We developed the technique of shot detection by cooperating with Shih et al. [2]. The video abstract extraction program extracts a n-second segment from each shot detected. It then combines these n-second segments into a *"Summary Video"*, which allows users to efficiently preview

the video content. Meanwhile, based on the shot detection results, the key frame extraction component extracts the appropriate frame from each shot to construct a JPEG format *"Key frame image"* file for static display.

5. The watermark appending module can embed an external image into every frame of a video file. A content provider can select a logo image and append it to a video to indicate ownership and discourage illegal use.

6. The format transformation module converts video data into different formats. For example, it can convert MPEG-2 files into MPEG-1, WMV, or RM formats. In addition, users can set up attributes for the output file, such as the frame size, bit rate and so on. Specifically, this module can generate a streaming file with a multi-bit rate format that can handle the various bandwidth of the Internet.

In Table 1, we list the time consumption for these modules. The testing environment is Windows XP with a P4 3.4G CPU and 1.5GB memory. The test data is 10 video files in MPEG2 format. The frame size of the files is 640*480, and the frame rate is 29.97 per second.

Table 1. The time cost of the video content analysis modules

Module	Time consumption rate (processing time / video duration)	Note
Caption recognition module	1.0 ~ 1.7	Depending on the number of captions
Caption appending module	0.60 ~ 0.63	
Voice recognition module	1.0 ~ 1.5	Depending on the number of video data
Shot detection module	0.7 ~ 1.0	Depending on the number of shots
Watermark appending module	0.65 ~ 0.70	
Format fransformation module	0.7-0.9	The output format is WMV with 352*240 frame size; the bit-rate is 364K.

After analyzing and processing, the video content is stored in DVAS using different formats, including text, image and video files. DVAS manages the data, which can be edited, searched, browsed, and used when required. We describe DVAS in detail in the next section.

4 DVAS Architecture and Implementation

4.1 The Components and Workflow of DVAS

DVAS preserves video metadata and digital video data. To enable the general public to browse and search video content online, DVAS comprises a metadata database, a

voice database, a video management and search system (VMSS), and a streaming server. Fig.2 illustrates the workflow of DVAS when responding to users' queries. The metadata database is responsible for saving injected metadata and the results of video caption recognition. The voice database is responsible for saving voice data obtained from the voice recognition module.

In DVAS, the VMSS provides capabilities for video management, such as metadata add/update/delete, and member authentication. It also provides a web query-interface and shows the query results obtained from the voice and metadata databases. To support real-time online viewing of videotapes, reduce the need for high network bandwidth, and protect intellectual property rights (i.e., prevent illegal copying), the DVAS utilizes a streaming server to play the video/audio content of videotapes.

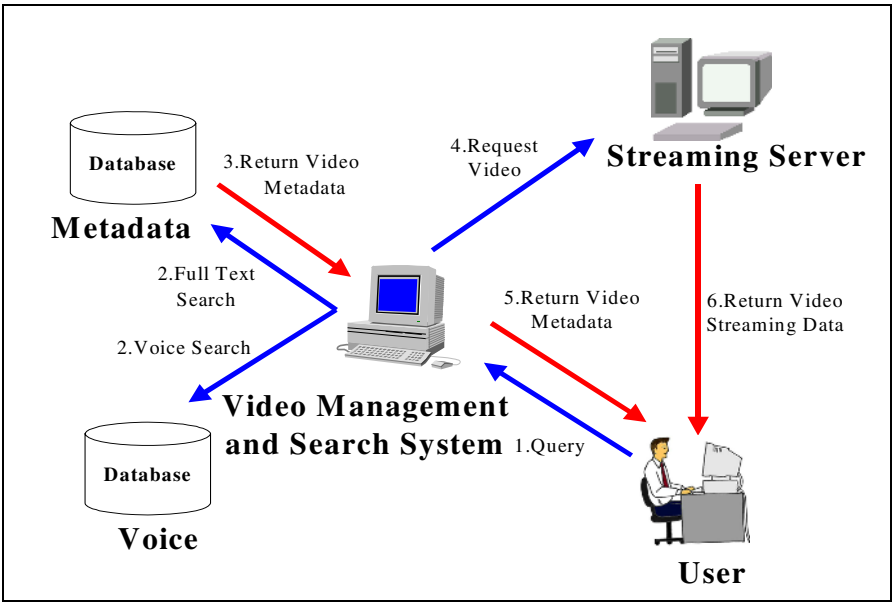

Fig. 2. The workflow of DVAS

4.2 The Implementation of DVAS

DVAS utilizes a 3-tier architecture: Apache and Tomcat Web programs serve as the application's server-tiers, and Oracle serves as the database-tier. We use the Linux Red Hat operating system for the server-tiers. The web pages were developed with JSP technology and Java Beans, and the video search engine was integrated with a streaming server for video/audio broadcasting. We use the Microsoft Media Server as the streaming server to publish WMV-format files. The hardware comprises two 1U servers with Intel Xeon processors to run VMSS and the streaming system. There is also a disk array that stores video abstracts and key frames as the total file size is 840 GB. Original videotapes are backed up with large tapes, because they are not accessed very often and the total file size is very large (over 8 TB).

4.3 Implementation Issues

4.3.1 Components for Video Content Analysis

We have already introduced several components for content analysis in Fig. 1. Each of these components is independent, so users can operate all or portion of them at the same time via several computers. In addition, all these components support batch operation, except the metadata injection system; thus users can deal with a large number of video files in one manipulation. As the components are independent, users can choose the minimum number of components to meet their requirements and easily reduce the processing time by using multiple computers. Furthermore, if the users have their own requirement for video content analysis, they can easily attach an external component to the system without modifying the original components.

4.3.2 Voice Data Search

In the DVAS, there are two kinds of metadata, one is purely text and the other is voice data. As we know, many words have a similar pronunciation, such as “two” and “too”. In this case, the voice recognition module may not choose the correct output results so that the recognition accuracy rate will decline. To solve this problem, our voice recognition module outputs the results in a format similar to phonetic symbol data, rather than as characters. When a user sends a query via the voice search function in VMSS, the query text will be automatically transformed into the above format and sent to the voice database for comparison.

4.3.3 Watermark Appending Module

The watermark appending module allows users to embed a logo image into every frame of a video. In this way, the user can claim ownership and prevent illegal usage of his/her video files. Although this sounds efficient, it raises two problems. The first is that the result cannot be reversed, once the process is finished, the original frames of the video are changed forever. The second problem is that processing takes a long time if the user wants to deal with a large number of video files. To solve the two problems, we use the FLV format, developed by Micromedia, as the streaming format in our new platform. Because this format has a multi-layer architecture, we can add a new image layer into video frames in real-time when members of the general public browse video content. In this way, we can provide rights protection for video files without the above problems.

5 Conclusions and Future Work

The Digital Museum of Taiwan's Social and Humanities Video Archive project was established three years ago. It has developed from digitizing original negatives to the formulation, entry, management of metadata, and video searching. A complete workflow of digital archive applications has been established and verified, and the project has yielded productive research results. Besides above archive project, we also provide these technologies to Digital Archives of Formosan Aborigines program [3] and Government Information Office, Republic of China (Taiwan) video archive system (in building). The two projects are important archive projects, and have rich video data (2000+ hours). Through our techniques, they can easily build a video management system and then provide the video data to general public.

We now indicate some future research directions.

1. Improve the workflow of automatic digitization. Currently, certain steps, such as the selection of key frames, are sometimes performed manually in order to select an appropriate image. In the future, we will integrate different methods to streamline manual processing, which will reduce errors and improve the system's overall performance.

2. Improve the technology for voice searching, as the accuracy rate of voice recognition in video files is not very good because of speakers' accents and background noise.

3. Integrate video copy detection technologies for digital rights protection. Currently, users often add non-visible watermarks to protect digital rights, but this is very costly in terms of computing time. Also, watermarks spoil parts of a frame, and their robustness against attack is not sufficient to guarantee security. Thus, we are developing technologies that will automatically find a video's feature information. Then, based on that information, we can compare two videos quickly. If the two video's features match, we may infer that the original video was probably pirated.

4. Open our sources to the public. We have already released the source code and execution files of the video format transformation and video watermark appending tools via the Open Source Software Foundry (OSSF) Web site [15]. By continuing to open our sources, we expect that more people will become involved, thereby promoting the development of the digital video archive.

5. Integrate content-based retrieval techniques. Content-based visual retrieval has received a great deal of attention from researchers in recent years. Users can use visual cues, such as color, texture, shape, and motion, to search perceptually similar video clips. Therefore, the integration of text and content-based retrieval would provide a more flexible way for users to process queries.

The complete Digital Museum of Taiwan's Social and Humanities Video Archive project has a vast amount of high quality content and employs several techniques to process and present it. Due to space limitations, we have only described the technologies and system architecture of DAVS. Other topics, including video content, e-learning systems, the design of metadata, and the development of information science technologies have not been discussed in this paper. In the future, we will continue in-depth research and development of these areas in order to construct a more advanced digital library.

Acknowledgements

This research was supported in part by the National Science Council of Taiwan under NSC Grants: NSC 90-2750-H-119-230, NSC 91-2422-H-119-0601, and NSC 92-2422-H-119-091. The authors wish to thank the members of the Digital Archive Architecture Laboratory (DAAL) for their assistance in building systems, and developing and integrating core techniques.

References

1. Carnegie Mellon University, “Informedia, digital video understanding research”, http://www.informedia.cs.cmu.edu/
2. C. C. Shih, H. Y. Mark Liao and H. R. Tyan “Shot Change Detection based on the Reynolds Transport Theorem”, *Proc. Second IEEE Pacific Rim Conference on Multimedia*, pp. 819-824, Beijing, China, October 2001.
3. Digital Archives of Formosan Aborigines program, http://www.aborigines.sinica.edu.tw/
4. Digital Museum of Taiwan's Social and Humanities Video Archive, http://www.sinica.edu.tw/~video/intro/intro-year10-e.html
5. Dulce Ponceleon, Arnon Amir, Savitha Srinivasan, Tanveer Syeda-Mahmood, and Dragutin Petkovic, “CueVideo: Automated Multimedia Indexing and Retrieval”, *ACM Multimedia '99* (Orlando, FL, Oct. 1999). p. 199.
6. European CHronicles On-line project, http://pc-erato2.iei.pi.cnr.it/echo/#
7. F. Chang, G. C. Chen, C. C. Lin, W. H. Lin “Caption analysis and recognition for building video indexing system“, *ACM Multimedia Systems Journal,* 10 (4), pp. 344-355, 2005.
8. Físchlár System, http://www.fischlar.dcu.ie/
9. G. Marchionini and G. Geisler, “The Open Video Digital Library”, D-Lib Magazine, Vol. 8, No. 12 (December 2002), http://www.open-video.org/
10. H. A. Wang, G. C. Chen, C. Y. Chiu, and Y.-C. Lin, “A case study on technologies of a digital video archive system”, *2005 International Conference on Digital Archive Technologies*, pp.101-113, Taipei, Taiwan, June 2005.
11. H. M. Wang, S. S. Cheng, and Y. C. Chen "The SoVideo Mandarin Chinese Broadcast News Retrieval System," *International Journal of Speech Technology*, 7 (2), pp. 189-202, April 2004.
12. Maryland Engineering Research Internship Teams, http://www.ece.umd.edu/MERIT
13. M. F. Huang, K. T. Chen and H. M. Wang, "Towards Retrieval of Video Archives based on The Speech Content," in Proc. *International Symposium on Chinese Spoken Language Processing (ISCSLP2002)*, Taipei, Aug 2002.
14. National Digital Archives Program, Taiwan, http://www.ndap.org.tw/index_en.php
15. Open Source Software Foundry, http://rt.openfoundry.org/Foundry/
16. Video Development Initiative, “ViDe User’s Guide: Dublin Core Application Profile for Digital Video”, http://www.vide.net/workgroups/videoaccess/resources/vide_dc_userguide_20010909.pdf