5th Pacific Rim Conference on Multimedia
(PCM2004), LNCS, vol.3332, pp.489-496

Mining Large-Scale Broadcast Video Archives towards Inter-Video Structuring

Norio Katayama[1], Hiroshi Mo[1], Ichiro Ide[2], and Shin'ichi Satoh[1]

[1] National Institute of Informatics
2-1-2 Hitotsubashi, Chiyoda-ku, Tokyo 101-8430, Japan
{katayama, mo, satoh}@nii.ac.jp
[2] Graduate School of Information Science, Nagoya University
1 Furo-cho, Chikusa-ku, Nagoya-shi, Aichi 464-8601, Japan
ide@is.nagoya-u.ac.jp

Abstract. The current computer technology enables us to build huge broadcast video archives which had been a future dream. Today, even the hard disk recorders on the market are capable of recording several hundred hours of broadcast video. It is naturally perceived that a huge amount of broadcast video would be a useful corpus for multimedia indexing and mining research. Based on this viewpoint, we designed and constructed a broadcast video archive system having sufficient capacity and functionality to serve as the testbed for indexing and mining research. The system can capture multiple channels (currently seven channels) all-day broadcast video streams simultaneously, up to 6000 hours, and program-specific broadcasts, currently a news program for more than three years so far. This paper discusses design and implementation issues of the video archive system and then introduces our research efforts utilizing the archives as huge multimedia corpora.

1 Introduction

The current computer technology enables us to build huge broadcast video archives which had been a future dream. Today, even the hard disk recorders on the market are capable of recording several hundred hours of broadcast video. It is naturally perceived that a huge amount of broadcast video would be a useful corpus for multimedia indexing and mining research. In addition, we should note that the speed of the technology progress is considerably fast. We observe that the capacity of hard disk drives grows ten times in every three years. Thus it might be possible, in five to ten years, to realize a personal set top box (STB) with terabytes to petabytes of disk drives, which can store thousands to millions of hours of videos. In such circumstances, it is crucial to establish component technologies of indexing and mining that are applicable to such huge multimedia corpora.

Based on this viewpoint, we designed and constructed a broadcast video archive system having sufficient capacity and functionality to serve as a testbed for multimedia indexing and multimedia mining research, which reflects a realistic scale for the future STB. In order to reflect the important nature of the huge broadcast video corpus, the system captures multiple channels simultaneously, 24 hours a day, as MPEG video files. The system also captures related text information including closed-caption text and electronic program guide (EPG) information. In realizing the system, instead of special hardware, we employ commodities for the components such as UNIX workstations, RAID disk arrays, and MPEG capture cards and closed-caption decoder cards installed in PCs. Captured data are managed by Oracle DBMS, and the system provides uniform view of access to the data for clients via Java JDBC API. We also developed the experimental video browser system which is intended to be used as the software platform of the system enabling rapid prototyping of video applications and video analysis software. This paper discusses design and implementation issues of the video archive system and then introduces our research efforts utilizing the archives as huge multimedia corpora.

2 Broadcast Video Archive System for Multimedia Research

2.1 Design Issues

In order to reflect the large-scale and dynamic nature of the broadcast video streams, the video archive system should meet several design issues. We first discuss desired functions of the system.

Desired video archives suited to multimedia indexing and mining research may need to keep capturing videos as a long period as possible, i.e., several years, while at the same time, they are required to capture as many streams (or channels) as possible for 24 hours a day. However, it is impossible to satisfy both requests at the same time, since in order to do this, required volume size of storage easily becomes prohibitively large. For important directions of multimedia indexing and mining research, there are mainly three demands for video archives making a compromise between the above two requests:

Diversity Capture of all-day video stream of as many channels as possible regardless of types of programs for a relatively small period, e.g., several days to several weeks.

Continuity Capture of particular programs broadcasted daily or weekly for a long period, e.g., several months to several years.

Autonomy Dynamic registration of captured data to archives reflecting everyday broadcasts.

Research efforts to handle continuous video streams (e.g., [1] realizes browsing of 24-hour broadcast videos) require diverse video archives. On the other hands, research attempts to analyze in-depth contents of specific types of programs (e.g., [2–5] concentrate on news, [6] on sports, [7] on cooking videos, etc.) may need continuous archives of specific types of programs. Therefore, desired video archives should satisfy both diversity and continuity. In realizing diverse and continuous video archives, daily or even hourly capture, registration, and management of video files are required. It obviously is impossible to manually construct the video archives having these characteristics. Automated capture, registration, and management are indispensable. Thus desired video archives may need to satisfy autonomy also, which tends to be lacking in the former attempts. CMU Informedia project [8] built the video archive system which satisfies autonomy for specific types of programs, namely, news and documentaries. But it does not meet diversity. We seek for all three demands at the same time for the video archive system.

2.2 Implementation Issues

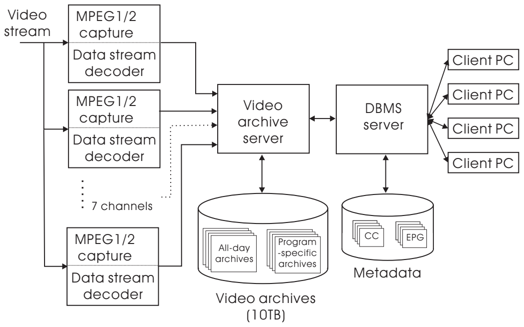


Fig. 1. Block Diagram of the Archive System

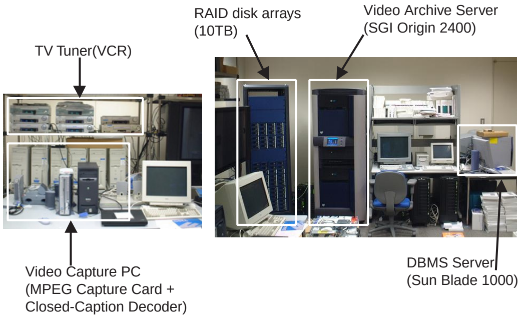

RAID disk arrays (10TB)
Video Archive Server (SGI Origin 2400)
TV Tuner(VCR)
Video Capture PC (MPEG Capture Card + Closed-Caption Decoder)
DBMS Server (Sun Blade 1000)

Fig. 2. Overview of the System

Recently, many commodities are becoming available which can be used as key components of the video archive system. However, there are no product which sufficiently satisfies our requirements. In implementing the video archive system, we decided to use commodities as components such as RAID disk arrays, MPEG capture cards, closed caption decoders, etc., and combine them on our own. Even though they are available as commodities, it is required to connect all these components in an integrated way, which is not trivial task, but rather challenging. In this section, we discuss implementation issues on the system by combining commodities.

Figure 1 shows the block diagram of the broadcast video archive system taking the demands discussed in the previous section into account. We use broadcast streams from seven channels (all terrestrial broadcasts available in Tokyo area) as sources of the video streams. The system can capture seven video streams simultaneously, and generate all-day video streams into one hour long MPEG files for each channel. In addition to video streams, the system also captures related text information including closed-caption (CC) text from data broadcasts, and electronic program guide (EPG) information from web. The system has an MPEG capture card (Canopus MVR-D2000) and a closed-caption decoder card (Systec Moji-Vision 550) installed in a PC for each channel, in total seven channels, to obtain required data.

Captured data are fed to the video archive server with RAID disk arrays. As the video archive server, SGI Origin 2400 is used, having 10TB RAID disk arrays connected through fiber channels. PCs are controlled by the server regularly, and to enable remote procedure calls (RPC) from the UNIX server to Windows PC, we use the Java remote method invocation (RMI) mechanism. The 10TB storage is split into 7TB for all-day streams (for diverse archives), and 3TB for particular programs (for continuous archives). Due to disk capacity limitations, the system keeps recent one month archives, about 6000 hours in MPEG-1 format in the 7TB storage. At the same time, the system captures one news program from NHK (the largest broadcast station in Japan) everyday, and keeps them persistently, so far for more than three years. For this type of archives, we capture video streams into MPEG-2 format to guarantee higher quality for the purpose of high-precision video processing such as face detection and motion analysis. The system captures these video streams in fully automated way, so it also satisfies autonomy. The system is also easily reconfigurable to capture other programs in addition to news.

Another component of the system is the DBMS server. We use Sun Blade 1000 for the hardware platform and Oracle for the DBMS software. The DBMS server regularly checks if there are any change on the video archives, and it autonomously reflects their changes to the database. The current implementation stores the following metadata in the DBMS server.

- Properties of MPEG video files (file path, broadcasted time, duration, channel number, etc.)
- Closed caption texts (each piece of closed caption text, broadcasted time, channel number, etc.)
- Electronic program guide (EPG) information (program guide text, start time of each program, channel number of the program, etc.)

Closed caption texts and EPG information are indexed by Oracle InterMedia Text so that they can be retrieved by the full-text search. These metadata provide the fundamental methods for locating and retrieving a particular video segment from the huge "video stream space".

By implementation techniques described here, the system is realized using commodities as components, and it works well satisfying requirements discussed in the previous section in an integrated way.

2.3 Software Platform for Prototyping

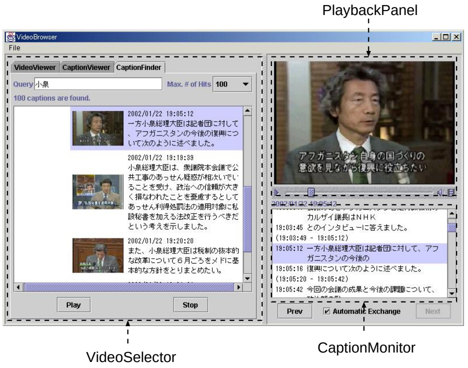
This video archive system is intended to be used by researchers for indexing and mining research. In principle, they can use this system by accessing MPEG files stored in the RAID disk arrays and their metadata in the Oracle database. However, it is not easy to develop research prototypes or experimental computer programs from scratch. Therefore, it is crucial to implement a software platform which provides handy application programming interface (API) for indexing and mining research. The software platform reduces not only the development cost but also the maintenance cost of the video archive system. As is common with other applications, the API encapsulates the internal configuration of the system and enables the system maintainer to evolve the configuration without interfering the end users of the system.

On designing the software platform, we focused on the following three requirements:

(1) Researchers should be able to develop their own computer programs easily which use the video files and their metadata stored in the system.
(2) The software platform should enable researchers to develop their programs on their own machines. This means that the software platform should provide the remote access to the video archive system.

(3) The software platform should work on various hardware platforms since the video archive system itself is an integration of various hardware components and since the researchers use a wide variety of hardware platforms including UNIX workstations and Windows PCs.

In order to fulfill these requirements, we employ Java programming language. Java is available on a wide variety of hardware platforms and it has a standard API for accessing remote DBMSs which is called JDBC (Java Database Connectivity). In addition, the combination of Java and the relational DBMS is so popular that we can take advantage of various information resources and development tools on the Web and on the market.



Query 小泉 Max. # of Hits 100

100 captions are found.

2002/01/22 19:05:12
一方小泉総理大臣は記者団に対して、アフガニスタンの今後の復興について次のように述べました。

2002/01/22 19:19:39
小泉総理大臣は、衆議院本会議で公共工事のあっせん疑惑が相次いでいることを受け、政治への信頼が大きく損なわれたことを憂慮するとしてあっせん利得処罰法の適用対象に私設秘書を加える法改正を行うべきだという考えを示しました。

2002/01/22 19:20:20
また、小泉総理大臣は税制の抜本的な改革について６月ごろをメドに基本的な方針をとりまとめたい。

2002/01/22 19:05:12

```
         カルザイ議長はＮＨＫ
19:03:45 とのインタビューに答えました。
(19:03:49 - 19:05:12)
19:05:12 一方小泉総理大臣は記者団に対して、アフ
         ガニスタンの今後の
19:05:16 復興について次のように述べました。
(19:05:20 - 19:05:42)
19:05:42 今回の会議の成果と今後の課題について、
```

☑ Automatic Exchange

Fig. 3. Video Browser

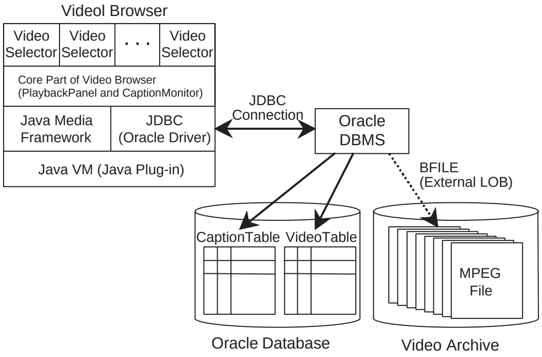


Fig. 4. Block Diagram of Video Browser

The API determines the usability of the software platform. In order to promote the rapid prototyping by researchers, we developed a Java applet which plays the role of the basis for prototyping. The applet is called “video browser” and has the following capabilities:

- Provides the fundamental GUI components for prototype systems.
- Enables researchers to accommodate their custom GUI components with providing the fundamental API for accessing the video archive.

The video browser consists of three types of components: PlaybackPanel, CaptionMonitor, and one or more VideoSelectors (Figure 3 and 4). PlaybackPanel plays a video stream, while CaptionMonitor displays text information synchronously to the video playback. VideoSelectors are GUI components for browsing/searching the video archive. Since they are placed on the tabbed pane, users can use multiple VideoSelectors interchangeably. By default, the video browser contains three VideoSelectors: VideoViewer (directory listing of video files), CaptionViewer (directory listing of textual information, i.e., closed caption and EPG), and CaptionSearch (full-text search of closed caption and EPG). In addition to these VideoSelectors, researchers can add their own VideoSelectors for testing their original indexing and mining methods.

As shown in Figure 4, the video browser accesses the MPEG files in the video archive through Oracle DBMS. Since Oracle DBMS has the mechanism to access external binary large objects, the video browser obtains both metadata and MPEG files only from JDBC connection. The obtained MPEG stream is fed to the player of Java Media Framework. Thus, the video browser requires only the JDBC connection to the DBMS server; it does not need other file access methods, such as network file system. This expands the coverage of the remote access to the video archive system.

3 Research Challenges in Mining Broadcast Video Corpora

We are now conducting multimedia indexing and mining research by taking advantage of the video archive system mentioned above. In this section, we introduce our recent efforts with presenting some lessons learned from using huge broadcast video corpora.

3.1 Inter-Video Structuring with Associating Video Segments

Broadcast video corpora potentially contains wide variety of information, which might be useful for developing advanced multimedia applications and evolving multimedia technology. While there exist various approaches for mining video corpora, one of the most fundamental approach is associating video segments to determine the inter-video structure among video streams. Association may cover any type of relationships, e.g., the appearance of the same person, recorded at the same location, dealing the same topic, etc. For example, in [9], broadcast videos of a daily news program are divided into video segments based on the topic boundary, and then the resultant segments are mutually associated based on the relevance between topics. By this approach, video segments are interweaved with the threads of topics. As mentioned above, our archive accumulates a daily news program for more than three years. Finding the topical relationship in a video corpus provides effective clues for inter-video structuring. Although the detection of inter-video relationships is the first step toward inter-video structuring, we believe that it is the essential and indispensable foothold.

3.2 Multimodality: Integrating Multimodal Information

Needless to say, multimodality is one of the most important property of broadcast video corpora. Video streams contains, textual(open and closed caption text), audio, and visual information. Since each type of information has different nature, it is important to integrate advantages of each type of information. For example, in [3], the face-name association is obtained based on the cooccurrence between a face image and a person name text. In [10], the key frame for the specific topic is detected based on the cooccurrence between frame images and topic threads. These methods successfully detect video-text relationships by integrating multimodal information. In [11], textual and visual information are used interchangeably in the video retrieval process. Since textual information is effective in expressing topic keywords and object names (persons, locations, etc.), it is used as a strongly restrictive filter. On the other hand, visual information is used for seeking useful or interesting video scenes from a bunch of video segments through browsing. This method reflects pros and cons of textual and visual information. The advantage of textual information is small processing cost and strong expressive power in restricting topics. On the other hand, the advantage of visual information is the efficiency in human perception, i.e., you can rapidly perceive the contents at a glance. As illustrated by above examples, the integration of multimodal information is an important strategy for utilizing huge multimedia corpora.

3.3 Mining Rare but Strong Inter-Video Relationships

Compared with textual information, it is quite difficult to use visual information for the clues to inter-video structuring. This is mainly because visual information is more subject to ambiguity due to the diversity of object appearances caused by the difference in pause, composition, lighting conditions, etc. However, if we focus on some particular type of visual information, it is possible to obtain strong inter-video relationships as mentioned below. Since this type of relationships can be obtained only for some specific type of visual information, they may rarely exist. However, when we have huge multimedia corpora, strong relationships are invaluable clues to inter-video structuring.

In [12], identical video segments are detected in a news program series, and it is reported that identical video segments are sometimes used for presenting particular topics or objects. In general, the frequency of identical video segments is not so high, but some symbolical and topical shots are often used repeatedly when some particular topic attracts public attention. The occurrence of identical shots are rather rare but once they are detected, they provide strong inter-video relationships. This strategy may not work with a small set of broadcast videos but it can be a strong tool for inter-video structuring when it is used with a large-scale broadcast video corpus.

Another example of mining rare relationships is face sequence matching with finding closest pairs[13]. Finding the similar face sequences is a basic function for finding the appearance of the same person. One of the difficulty in finding the similar face sequences is the diversity of recording conditions, e.g., lighting, pause, facial expression, etc. Although the conditions may vary for each face sequence, two face sequences may happen to contain two face image pair that are very close to each other. Based on this viewpoint, we developed a face sequence matching method that evaluates the similarity of face sequences by the similarity between the closest pair. Although the search for the closest pair is computation intensive, this method gives better precision than the method which evaluates the similarity of face sequences by comparing best-frontal face image pairs.

3.4 Online Management of Large-Scale Video Corpora

From the implementation aspect, an important nature of the broadcast video archive system is that it is an online system and archives grow continuously. In order to reflect the latest broadcast information, indexing and mining methods must be adapted to the online processing. In addition, as the storage amount and the number of recording channels increase, the requirement for the online processing should be more crucial. At this moment, we do not have concrete solution to this problem but we expect that the database technology would play an important role in extending the scalability.

4 Conclusions

We designed and constructed a broadcast video archive system having sufficient capacity and functionality to serve as a testbed for multimedia indexing and multimedia mining research. Three demands were pointed out for desired video archive system, i.e., diversity, continuity, and autonomy. Actual implementation of the system satisfying these demands is shown using commodities as its key components, such as UNIX workstations, RAID disk arrays, and MPEG capture cards and closed-caption decoder cards installed in PCs. Then to develop video applications and video analysis software, the software platform of the system is also introduced for rapid prototyping of video applications and video analysis software. By using the constructed testbed, we are now tackling multimedia indexing and mining techniques towards inter-video structuring in huge multimedia corpora.

References

1. Yukinobu Taniguchi, Akihito Akutsu, Yoshinobu Tonomura, and Hiroshi Hamada, “An intuitive and efficient access interface to real-time incoming video based on automatic indexing,” in *Proc. of ACM Multimedia*, 1995, pp. 25–33.
2. Yuichi Nakamura and Takeo Kanade, “Semantic analysis for video contents extraction — spotting by association in news video,” in *Proc. of ACM Multimedia 97*, 1997.
3. Shin’ichi Satoh, Yuichi Nakamura, and Takeo Kanade, “Name-It: Naming and detecting faces in news videos,” *IEEE MultiMedia*, vol. 6, no. 1, pp. 22–35, January-March (Spring) 1999.
4. Michael G. Christel, “Visual digests for news video libraries,” in *Proc. of ACM Multimedia*, 1999, pp. 303–311.
5. Xinbo Gao and Xaioou Tang, “Unsupervised and model-free news video segmentation,” in *Proc. of IEEE Workshop on Content-Based Access of Image and Video Libraries*, 2001, pp. 58–64.
6. N. Babaguchi, S. Sasamori, T. Kitahashi, and R. Jain, “Detecting events from continuous media by intermodal collaboration and knowledge use,” in *Proc. of International Conference on Multimedia Computing and Systems (ICMCS)*, 1999, pp. 782–786.
7. Reiko Hamada, Shin’ichi Satoh, Shuichi Sakai, and Hidehiko Tanaka, “detection of important segments in cooking videos,” in *Proc. of IEEE Workshop on Content-Based Access of Image and Video Libraries*, 2001, pp. 118–123.
8. Howard D. Wactlar, Michael G. Christel, Yihong Gong, and Alexander G. Hauptmann, “Lessons learned from building a terabyte digital video library,” *IEEE Computer*, vol. 32, no. 2, pp. 66–73, 1999.
9. I. Ide, H. Mo, N. Katayama, S. Satoh, “Topic threading for structuring a large-scale news video archive”, *Int. Conf. on Image and Video Retrieval (CIVR2004)*, LNCS vol.3115, 2004, pp.123–131.
10. H. Mo, F. Yamagishi, I. Ide, N. Katayama, S. Satoh, and M. Sakauchi, “Key Image Extraction from a News Video Archive for Visualizing its Semantic Structure,” *PCM 2004* (to appear).
11. C. Yu, H. Mo, N. Katayama, S. Satoh, and S. Asano, “Semantic Retrieval in a Large-Scale Video Database by Using both Image and Text Feature,” *PCM 2004* (to appear).
12. F. Yamagishi, S. Satoh, and M. Sakauchi, “A News Video Browser Using Identical Video Segment Detection,” *PCM 2004* (to appear).
13. S. Satoh and N. Katayama, “An Efficient Implementation and Evaluation of Robust Face Sequence Matching,” *Proc. of ICIAP99*, 1999, pp. 266–271.