Proceedings of the 5th WSEAS Int. Conf. on MULTIMEDIA, INTERNET AND VIDEO TECHNOLOGIES, Corfu, Greece, August 17-19, 2005 (pp108-113)

Storing Video Data with High Speed Access in Hierarchical Storage Systems

Shigeo Goto†, Tadao Nakamura†, Masatsugu Hashimoto‡ and Norio Izumi‡

† *Graduate School of Information Sciences, Tohoku University*
6-6-01 Aramaki Aza Aoba, Aoba-ku, Sendai, 980-8579, JAPAN

‡ *SGI Japan, Ltd.*
Yebisu Garden Place Tower 31F, 4-20-3, Ebisu Shibuya-ku, Tokyo 150-6031 JAPAN

Abstract: - This paper describes a high speed media access from a hierarchical storage system consisting of disks and auxiliary storage media. Generally, the storage system to store video data has slow access time. Especially, if the auxiliary storage media consists of tapes, the storage system has too long access time due to the handling time between disks and tapes even though they are controlled under the storage system. In this paper, we proposed to replace the tapes with DVD, so that the access time of the storage system consisted of disks and DVDs are shorter than that in the disks and tapes. The key of this paper is to further reduce the DVD access time by means of data compression of video data on the DVD with almost the same access time of the disk. As a result, this storage system becomes a large volume disk even though this consists of the disk and the DVD.

Key-Words: -MPEG-4, Hierarchical Storage System, DVD-RAM Library

1 INTRODUCTION

Storing digital video data in storage systems is very important in terms of avoiding the decay of video tape and promoting the mutual use of video data. Especially, NHK and other broadcasting companies have started the use of this technology to store video data in tapes into the storage systems [1][2]. As tapes have large volume in the long span tracks, exchanging the data of tapes into data in the digital storage system needs lots of volume of storage systems. Besides, the access from editors takes a lot of time in terms of data transfer.

Here, there are naturally two problems of how to reduce the cost of large-scale storage system and how to get shorter access time in video data viewing. While dealing with archives by means of large-scale storage systems, these problems are strongly emphasized and become seriously recognized.

Storing huge data even in large-scale storage systems induces large cost because of the large volume of storage capacity and their complicated control systems. If these storage systems are connected through networks, the throughput and time delays requests us to provide the efficient protocols and physical conditions [3][4]. The cost of storage systems consists of the aforementioned factors.

The access time of these storage systems is another evaluation factor in addition to data transfer time delay [5]. These two factors are related to the response time of video data viewing.

Section 2 shows the outline of our research objectives and describes the problems of storing huge volume for video data. Section 3 explains the design of large-scale storage system with the solutions to the above problems. We implement a storage system with the proposed design method and further progress in evaluation of the system. Section 5 summarizes these design process.

2 Media Data Storing

In order to preserve high quality video and provide them to viewers, broadcasting companies, postproductions store and control video data by using digital video tape recorders (DVTRs). Table 1.1 shows the various kinds of DVTRs, where they are many kinds of DVTRs such as D1, D2, D3, D5, DV, DVCPRO, DVCPRO50, Digital BETA-CAM and etc.

Currently, there is a high need to store and manage video data stored in DVTRs in terms of avoiding the decay of video tape and mutual uses of video data. However, video data involves huge volume of data and consumers a lot of time in terms of data transfer.

2.1 Data Capacity

Based on IRU-R BT.601, we have the capacity of video data in the NTSC method. Every line of a 525/60 TV picture has 720 luminance (Y) samples and 360 each of two chrominance samples (Cr and Cb), making a total of 1,440 samples per line. For analogue signals, there are 487 active lines (480 for ATSC digital) and therefore 1,440 x 487 = 701,280 pixels per picture (691,200). With

Table 1.1 Digital Video Tape Recorder

VTR	Signal	Y:Cr:Cb	Sampling	Compression	(horizontal pixels) × (Active lines) Y	Cr, Cb	Tape Width
D-1	Component	4:2:2	8bit	-	720×500	360×500	3/4inch
D-2	Composite	-	8bit	-	768×510	-	3/4inch
D-3	Composite	-	8bit	-	768×510	-	1/2inch
D-5	Component	4:2:2	10bit	-	720×510	360×510	1/2inch
Digital BETA CAM	Component	4:2:2	10bit	1/2	720×512	360×512	1/2inch
DVCPRO50	Component	4:2:2	8bit	50Mpbs	720×480	360×480	1/4inch
DVCPRO	Component	4:1:1	8bit	1/5	720×480	180×480	1/4inch
DVCAM	Component	4:1:1	8bit	1/5	720×480	180×480	1/4inch
DV	Component	4:1:1	8bit	1/5	720×480	180×480	1/4inch
D-VHS	Component	4:2:0	8bit	MPEG2	720×480	360×240	1/2inch
DVD-Video	Component	4:2:0	8bit	MPEG2	720×480	360×240	-

each pixel sampled at 8-bit resolution, this format creates 5,610,240 bits, or 701.3 kilobytes (KB). At 30 frames per second, this format creates a total of 21,039 KB, or 21 megabytes (MB) per second. Note that both 625 and 525 line systems require approximately the same amount of storage for a given time 21 MB for every second. To store 76 gigabytes (GB), it takes one hour.

Thus, large-scale storage systems are necessary for storing video data that broadcasting companies and postproductions holds. As the storage cost of video data increases proportionally to the scale of storage systems, the approach to reduce video data capacity is to keep the storage cost lower. However, when video data are required to replay by multiple viewing systems, broadband networks need to transfer video data stably to multiple viewing systems. Implementing broadband networks require a lot of cost, therefore in order to keep the cost of network implementation lower, we are forced to consider the reduction of the capacity for video data.

As described above, the storage of huge video data needs large-scale storages systems and broadband networks with large cost for these implementation. To solve this problem, we must think of data format, data compression, and also the management on video data.

2.2 Storage Cost

The realization of mutual uses for video data make view instantly video primitive apart from old linear editing system by tapes, digital processing by digital non linear editing system, and improvement of these convenience, thus reduction of creation time. Therefore, storages systems with lower cost are needed. Table 2.1 shows the properties of storage medias.

In storage media prices, ultra320 Hard Disk Drives (HDD) per 1 MB is 18 times as high as the price of LTO. If we use half a million tapes of 60 minutes stored in DVTR of D1, the storage capacity of video data is 40 petabytes (PB). This amount is stored in HDDs or tapes based on these conditions we calculate the storage cost of these to HDDs and tapes.

In case of storing video data are of 40 PB on ultra320 as for SCSI HDDs, it costs 1152 hundred million yen. On the other hand, in the case of LTO, it costs 64 hundred million yen, therefore broadcasting companies use lots of tapes as the asset with no less than hundreds of thousand of tapes. If these video data are stored in storage systems, such petabytes costs huge amount with them. Depending on sorts of storage media costs vary in a wide range therefore viewing frequency video data and fast data transfer request clear specification of storage systems and wide selection of them.

2.3 Viewing Response

Viewing response is of response from the request of viewing video data to show of video on the monitor. In order to view video data, video data stored in storage systems should be take out on the viewing systems. In this case transfer rate of the storage systems affects the viewing response. When video data of D1 is viewed in real-time, we need data transfer rate of about 21 MB/s. If data transfer rate of the storages system cannot reach the transfer rate available for viewing in real-time, we view video data through the transfer to higher storage system.

Table 2.1 Storage media

Group	HDD	Optical Disc			Magnetic Tape		
Type	Ultra320 SCSI	MO	CD-R	DVD-RAM	DAT (DDS-3)	DLT (DLT7000)	LTO (Ultrium1)
Size	72GB	640MB	700MB	9.4GB	12GB	35GB	100GB
Unit Cost	\2.88 /MB	\1.22/MB	\0.29/MB	\0.37/MB	\0.16/MB	\0.28/MB	\0.15/MB
Data Transmission Rate	40MB/Sec	5MB/Sec	3.6MB/Sec	5MB/Sec	1MB/Sec	10MB/Sec	15MB/Sec

Proceedings of the 5th WSEAS Int. Conf. on MULTIMEDIA, INTERNET AND VIDEO TECHNOLOGIES, Corfu, Greece, August 17-19, 2005 (pp108-113)

The in case of lower data transfer rate of the storage system the view response time is poor due to data transfer of video data. According to data transfer rate shown in Table 2 of storage systems viewing of D1 video data possible in real-time is on HDDs. The other storage systems need data transfer into higher storage systems to make a view.

3 System Design and Implementation

We show here the design and implementation of storage system solving the response problem with storage cost.

3.1 Hierarchy of Video Data

Storing huge video data needs large-scale storage systems and broadband networks, therefore in order to reduce video data, we introduce a system for managing hierarchical video data.

Hierarchical video data is defined to be managed master video data and viewing video data hierarchically. Master video data is the original video data of broadcasting video. Viewing video data is the one that is compressed from the master video data. The figure of merit for hierarchical video data is available for large saving of storage systems capacitance limited video data of viewing stored storage systems. As viewing video data has a much small capacity than master video data, the traffic of networks becomes smaller. Therefore, under the lower cost of storage systems, high-speed response viewing is available.

3.2 Hierarchy of Storage Systems

Broadcasting companies store dozens of tapes a day. Therefore basically the video data storage systems must have large-scale capacity of storage to keep storage cost lower let us introduce hierarchical storage systems capable of consisting cheaper storage systems. As hierarchical storages systems DVD-RAMs superior in viewing video data are used. DVD-RAMs has outstanding random access capability and has no damage of storage media thanks to non-touchable I/O mechanicals. Further these I/O units are very simple. From these features, we can conclude that DVD-RAMs are superior to tapes.

3.3 Video Format

Video format of viewing video data is in a view quality capable of recognizing the contents at viewing and is obliged to keep data capacity lower. For long terms from these points of view we must use video format of MPEG-4 that is available for video viewing with less data quantity. As MPEG -4 is MPEG (Moving Picture Expert Group) international standard, we can support long for video format of specify venders.

3.4 Transfer Format Video Data

It is necessary for video data to be transferd from storage systems replay systems at viewing video data. While regarding the response viewing video data as important, transfer format of video data should be a streaming method. A streaming method is one in which successfully replay is performed as soon as video data reaches replay systems .A streaming method do not have the necessity to wait for all transfers of all video data because successive replay performed as soon as data arrive. Therefore, viewing has quick response.

3.5 Data Cache

Hierarchical storage systems are superior to hard disks in costs since we can construct large-scale storage systems by using cheaper media. However, hierarchical storage systems have the property of large delayed I/O of data such as slow read and write processing of storage media themselves and control time for multiple storage media. Then there is the problem of response between requirement of viewing video data and practical monitoring of them. In order to solve the problem of data transfer rate in such a hierarchical storage systems, let us introduce the data cache concept. Data caching of video data is required by replay systems constructing data cache systems with high-speed disks.

3.6 Systems Implementation

Fig 3.1 shows the whole system that consists of the contents for distribution server systems, the contents for registration server systems, hierarchical storage systems, registration terminals and replay terminals.

Contents distribution server systems are with MediaBase by Kassenna corp. MediaBase is composed of streaming server by practically distributing the video data, Mediabase cache caching the video data and MediaBase DB managing metadata and distribution logs. Registration server systems are the ones that manage video data and their meta-data. Hierarchical storage systems perform I/Os between high-speed storage media and library systems. This constitution consists of AMASS to manage hierarchical storage systems, high-speed storage media AMSS cache and the DVD library systems. Registration terminals perform the registration of video data and meta-data. Replay terminals works by taking out and replaying of the stored video data.

Proceedings of the 5th WSEAS Int. Conf. on MULTIMEDIA, INTERNET AND VIDEO TECHNOLOGIES, Corfu, Greece, August 17-19, 2005 (pp108-113)

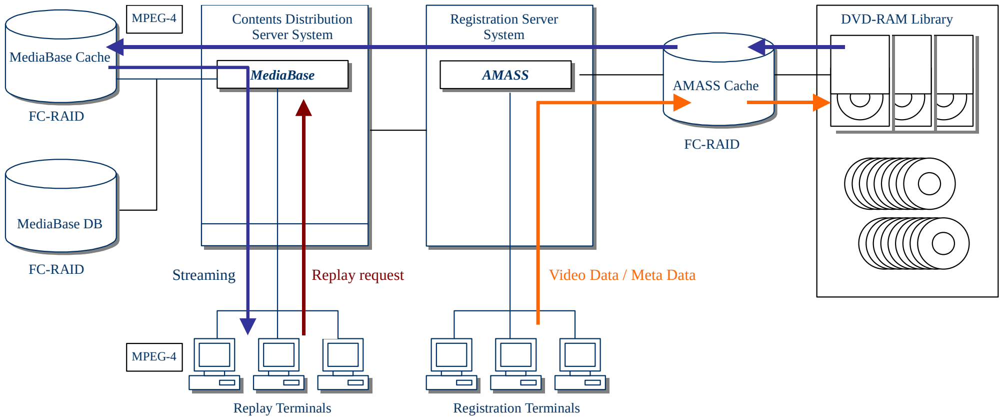

Fig. 3.1 Systems overview

4 Performance Evaluations

In this section, we will evaluate the hierarchical storage systems in terms of storage costs and viewing response times.

4.1 Evaluation of Storage Costs

Evaluation of storage costs consists of individual video format and storage costs hierarchical storage systems. The former means the comparison of storage costs after calculation of video data capacitance per video format. In order to calculate video data capacitance, we must calculate data capacitance of digitalized video data for each video format. In practical systems, video data controlling is hierarchical structured, dividing video data into original ones and viewing video ones. Therefore, video data stored in this storage systems means viewing video data. We are going to evaluate the storage costs in transforming the viewing data into digital data with MPEG-2, MPEG-4, etc.

The algorism of evaluations is shown below;

- Evaluation cycle: Calculating data capacity at transformation into video data for hundred thousand of 60 minutes video data.
- Video format is by MPEG-4, MPEG-2, DVCPRO50 or non-compression.
- Storage of video data for viewing produced in the above video format.
- Storage medias are MO, CD-R, DVD-RAM, DAT, DLT or HDD.
- Media costs according to the Table 2.1 in section 2.

These results are shown in Table4.1 for video data capacity of cache video format and storage cost for each storage media.

Let us calculate the data capacity for each video format. The data capacity required for data storage per hundred thousand video data are 22,5 terabytes (TB) for MPEG-4, 360 TB for MPEG-2, 2250 TB for DVCPRO50 and 7560 TB for non-compression, which indicates the necessity of large-scale storage systems. In comparison of these volumes, we normalize these volumes by data capacitance, 22.5 TB for MPEG-4. Then, we get 16 times for MPEG-2, 100 times for DVPRO50 and 336 times for non-compression. This means the increase data capacitance proportional to bit rates of video data compression rates. Video data to be stored in the system is video data for viewing, which means keeping high quality enough to viewing. MPEG-4 is used in viewing from cell-phones to STB, IP-Multicast TV or monitor cameras. From these practical efforts MPEG-4 is regarded as high quality level for viewing uses. Therefore, we could think that MPEG-4 is suitable for video format satisfying video quality of viewing. Using MPEG-4 can reduces the data capacity of storage systems much larger than any other video format. Therefore it costs smaller than video data capacitance in keeping the storage cost as small as possible.

Table4.1 Data capacity of each video format and storage cost of each storage media

	Data Capacity (60 minutes and one hundred thousand)	Media cost (yen in millions)					
		MO	CD-R	DVD-RAM	DAT	DLT	HDD
MPEG-4 (500Kbps)	22.5TB	27.5	6.5	8.3	3.6	6.3	64.8
MPEG-2 (8Mbps)	360TB	439.2	104.4	133.2	57.6	100.8	1036.8
DVCPRO50 (50Mbps)	2250TB	2745.0	652.5	832.5	360.0	630.0	6480.0
Non-compression (D1)	7560TB	9223.2	2192.4	2797.2	1209.6	2116.8	21772.8

Table4.2. Specifications of DVD-RAM and TAPE storage system

	DVD-RAM	TAPE [*1]
System size	1950(H) x 900(W) x 1400(D)	2349(H) x 3250 (R)
Media size	1.2(H) × 120 (R)	25.88(H) x 109(W) x 125(D)
Number of media (max.)	4620	6000
Media capacity	9.4GB	20GB
Drive size	41.3(H) x 146(W) x 196(D)	160(H) x 230(W) x 480(D)
Data transfer rate	10MB/sec	10MB/sec
Number of drive (max.)	88	40
Media life	30 year	10 year
maximum storage capacity	43.428TB	120TB
Library system space	2457000000 mm3	19486759268 mm3
Storage capacity per (per 1mm cubic meter)	17676 Byte/mm3	6158 Byte/mm3

*1: Stragetek 9840

Next, let us consider the storage costs of each hierarchal storage system with different medias. Each storage media cost in the Table 4.1 shows the storage cost at storing each storage media storage systems of with data capacitance each video format. Calculation of storage costs use unit cost per 1 MB in the Table in 2-Section. In considering MPEG-4, storage costs are 3.6 yen for DAT, 6.3 yen for DLT, 6.5 yen for CD-R, 8.3 yen for DVD-RAM, 27.5 yen for MO and 64.8 yen for HDD respectively. If we see these costs according to media, they need storage costs in the order of tape media, optical media, and HDD media. Noting the media with highest storage density among these media, DVD-RAM has the highest density in optical media and DLT is the highest in tapes. If we compare these costs adding HDD to these to media, DVD-RAM is 1.3 times the cost of DLT and HDD is 10.3 times the cost of DLT. Thus, constituting the hierarchical storage system with storage media cheaper than HDD, we can reduce the storage cost of 90.3% for DLT and 87.2% for DVD-RAM compared to the storage systems with HDD.

Next, let us evaluate the storage capacitance per unit capacitance for the hierarchical storage systems consisting of tape media and optical media. In tape media and optical media storage systems, we must calculate storage density of stage media to cabinet of stage systems. The expression of density evaluation is given by the following equation.

$$\text{Storage Density per Unit Capacitance} = \frac{\text{Media Capacitance} \times \text{Maximum Numbers of Media}}{\text{Capacitance of a Storage System}}$$

The result is shown in Table 4.2. From Table 4.2, storage capacitance per 1 mm meter square capacitance is 17676 for DVD-RAM and 6158 for tapes. Therefore, when comparing storage capacitance per unit volume, DVD-RAM has 2.9 times the storage capacity of tapes. Therefore, we can reduce installation space and keep the cost low.

4.2 Evaluation of Viewing Response of Video Data

Fig 4.1 shows the flow diagram of viewing video data. We provide for physical distance between the replay terminal and video data media that are located in different positions A, B, C and D in order to measure transfer delays of video data. Point A means streaming server with video data cached. Point B has a hierarchical storage system with video data in the data cache. Point C has DVD-RAM drive with video data. Point D has DVD-RAM library system with video data. Results of measures are shown in Table 4.3.

The distance B-C emphasizes the effectiveness of the data cache for MediaBase. When the data cache is in the hit status, transfer delay is reduced by 28.2s. The distance C-B and the distance D-C show the effectiveness of AMASS's data caches. In the distance C-B where DVD-RAM with its drive has video data, we can reduced the transfer delay by 2.6s when AMASS's data cache is in the hit status. Next, the distance D-C means the DVD-RAM library with video data, where we have the reduction of transfer delay by 48.2s at cache hit of AMSS. Lastly, the distance D-A has 73.8s of reductions in transfer delay at media base data cache in the hit status.

Next, the hit rate data cache in a streaming server for two weeks we measured the number of hit in this cache viewing video data using by implementing the proposed system. The result is shown in Table 4.4. As results of measures, we have got the number of viewing video data, 288.9 times, where the number of hits in the data caches is 211.1 times, the number of accesses to hierarchical storage system is 77.8 times. From these results, the frequency of hits to the data cache is 71.4% and the one to hierarchical storage system is 28.6%. Hence, about 70% of video data at transferred rapidly, which manage the delay of data transfer in the hierarchical storage system. At the same time, the hit frequency data cache keeps the network traffic between the streaming server and hierarchical storage system lower.

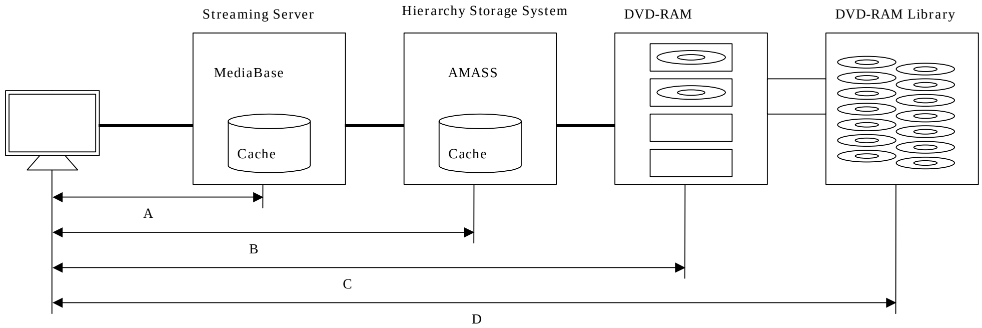

Fig.4.1 Flow diagram of viewing video data

Table 4.3 Response time of viewing video data

Data Transfer Distance \ Number of Times	1	2	3	4	5	Average	B-A	C-B	D-C
A	15	12	16	16	12	14.2	28.2		
B	34	33	45	35	52	39.8		2.6	
C	42	35	43	46	46	42.4			48.2
D	84	87	85	85	99	88.0			

5 Conclusions

We have proposed to implement a hierarchical storage system in order to solve the problem of cost of video data and response at viewing. In order to decrease storage costs we have implemented the hierarchical storage system using DVD-RAMs for decreasing the data capacity in MPEG-4. On the other hand, for the problem of response for video data viewing, we have solved the delay of data transfer in the hierarchical storage system by using data cache.

Actually, the hierarchical storage system described here is currently used in some broadcasting company as viewing video data storage systems. Viewing video data is stored by slow degrees in the hierarchical storage system and finally we are going to store viewing video data with hundreds of thousand of videos.

One of the future problems is to optimize the capacity of data caches and the policy of caches. While operating the hierarchical storage system, we are going to improve this system with a much better response in viewing video data compared to conventional methods.

Table 4.4 Hit Rate Data Cache in Streaming Server

No.	Total number of accesses	Total number of Data Cache Hit		Total number of accesses for hierarchical storage	
1	96	61	63.5%	35	36.5%
2	362	257	71.0%	105	29.0%
3	367	269	73.3%	98	26.7%
4	417	323	77.5%	94	22.5%
5	320	236	73.8%	84	26.3%
6	574	444	77.4%	130	22.6%
7	83	56	67.5%	27	32.5%
8	28	20	71.4%	8	28.6%
9	362	290	80.1%	72	19.9%
10	313	221	70.6%	92	29.4%
11	267	178	66.7%	89	33.3%
12	326	203	62.3%	123	37.7%
13	444	339	76.4%	105	23.6%
14	86	59	68.6%	27	31.4%
Average	288.9	211.1	71.4%	77.8	28.6%

References:

[1] NHK, NHK Archives, http://www.nhk.or.jp/nhk-archives/

[2] A.Mori, T.Nimi, S.Otani, S.Masui, M.Yamamoto, Y.Tokoro, M.Yoshimura, “System Overview of NHK Archives”, BROADCAST ENGINEERING 2003.2, pp.59-77, 2003.

[3] Chan,S.-H.G., Tobagi,F.A., "Modeling and dimensioning hierarchical storage systems for low-delay video services",Computers, IEEE Transactions, Vol.52, Issue 7, pp.907-919, Jul 2003.

[4] Li Qingguo, Sun Songlin, Cai Anni, Sun Jing-ao, "Hierarchical storage design using queueing model in VOD system", Info-tech and Info-net, ICII 2001-Beijing, International Conferences, Vol.2, pp.786-790, Nov 2001.

[5] Ogawa,S., Kamimura,K., Kato,T., Uehara,T., Okuda,H., "Performance analysis of hierarchical storage management systems for video retrieval system", Consumer Electronics, ICCE. International Conference, pp.328-329, June 2001.