

動画サーバの高速多重処理技術

西村 一敏

丸山 充

阪本 秀樹

中野 博隆

NTTヒューマンインタフェース研究所

ビデオオンデマンドに用いられる動画サーバでは、番組のコピーなしでの多重読取り技術、その多重度を維持できる特殊再生技術、および番組要求・特殊再生要求時の高速応答技術が重要な課題である。本報告では、複数磁気ディスク装置を用いる並列読取り方式の性能を理論解析している。並列度 n では待ち時間をシングルディスクの場合の n 倍にすれば、読取り多重度を n 倍にできる。また、高多重読取り方式とその特殊再生法・待ち時間短縮法を提案し、試作実験機により検証している。本方式は、バッファメモリと待ち時間の増大なしに、読取り多重度を n 倍に改善できる。磁気ディスク装置 4 台を用いて、1.5Mbps 動画の 52 多重読取りが達成された。

HIGHSPEED AND HIGH-MULTIPLE PROCESSING TECHNIQUES FOR VIDEO SERVERS

Kazutoshi Nishimura

Mitsuru Maruyama

Hideki Sakamoto

Hiroataka Nakano

NTT Human Interface Laboratories

3-9-11, Midori-Cho, Musashino-Shi, Tokyo, 180 Japan

Key techniques of video servers are multiple readout without any copy of video programs, trick play compatible with the multiple redout, and quick response to playback requests. This paper describes the theoretical performance of parallel-disk readout. Parallel readout from n magnetic disk drives increases multiple-readout number to n times together with the n -times increase of waiting time. A new technique is proposed and experimentally verified. This technique produces the same multiple-readout improvement without the increase of buffer memory and waiting time. The fifty-two simultaneous readout of 1.5-Mbps video has been achieved using four drives.

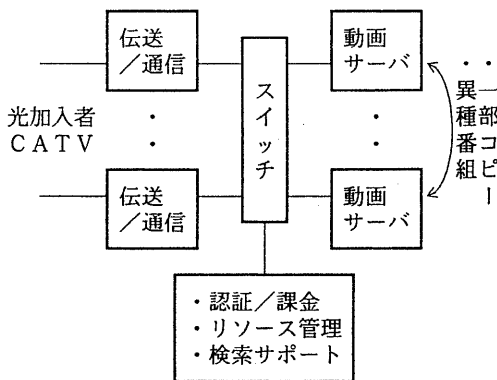


図1 VODセンタの構成

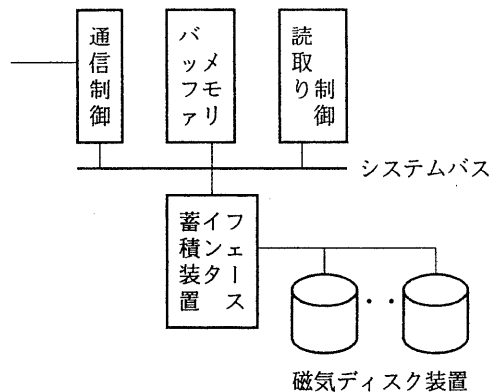


図2 動画サーバの基本構成

1. まえがき

ビデオオンデマンド（VOD）は、通信網を介してセンタから個々の要求動画番組を端末に提供するサービスである。近年、通信網のデジタル化、動画符号化（圧縮）技術の進展、および磁気ディスク装置の高性能化を背景に、全デジタル方式が提案されている[1]。

動画サーバの読取り方式に関しては、文献[1]で、ニアオンデマンドの待ち時間を数分に短縮する高性能磁気ディスク装置の繰返し読取り方式が提案されている。一方、筆者らは、磁気ディスク装置を時分割で共同利用し、数秒以下の待ち時間を実現する多重読取り方式を考案した[2]。同様の提案は文献[3]でもなされている。

本報告では、理論解析により、従来の多重読取り方式の性能を明らかにするとともに、新方式である高多重読取り方式とその早送り等の特殊再生法、待ち時間短縮法について述べる。

2. VODセンタの構成

図1に地域規模VODサービスのセンタ構成を示す。各家庭の端末とセンタは、光加入者やCATVなどのネットワークで接続される。同時にサービスを提供すべき端末数が多重読取り性能を超えるので、複数台の動画サーバが必要になる。利用頻度

の大きな人気番組を別にすれば、経済性の観点から、各動画サーバには異なった動画番組が蓄積される。従って、特定の番組を蓄積した動画サーバの読取りチャンネルと端末への伝送チャンネルとを接続するために、スイッチが必要となる。

このスイッチとして、筆者らは、MISS (Multi-media Integrated Swiching System) を提案し、試作検証を済ませた[4]。クロスバススイッチとスケジューリング転送により、総合で実効 2.4Gbps の効率の良いデータ転送を実現している。

動画サーバに関しては、番組をコピーすることなしに1台で同時にサービスできる端末数、すなわち読取りと通信の多重度を向上することが大きな技術課題となっている。通信については、動画伝送のレート制御とエラー補償が必要であり、動画サーバ以外で分担させることもある。

3. 多重読取り方式

3.1 多重読取り方式の原理

図2に動画サーバの基本構成を示す。磁気ディスク装置上の動画番組は、読取り制御部の制御の下で、蓄積装置インタフェース部を介してバッファメモリに読取られる。さらに、バッファメモリから通信制御部を介して、端末へ転送される。動画番組の多重読取りは、磁気ディスク装置の読取

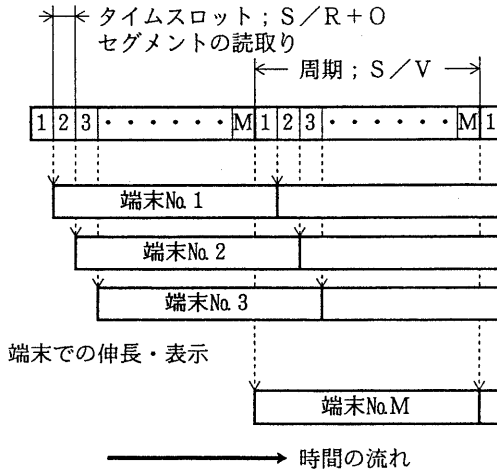


図3 多重読取りの動作原理

り速度と動画番組のビットレートとの速度差を利用して行う。その動作原理を図3に示す。周期的に繰返すタイムスロット1～Mは、それぞれ端末1～Mのために使用される。

例えば、符号化された動画番組を1秒分毎のセグメントに分割して、磁気ディスク装置に蓄積する。0.1秒のタイムスロットで、1つのセグメントへアクセスし、かつ読取りデータを端末のバッファメモリへ転送する。端末では、これを1秒かけて伸長しつつ表示する。その終了前に次の周期のタイムスロットで、引続くセグメントのデータを補給する。この間には他の9端末に対しても同様に処理できるので、10端末に対する個々の要求番組の同時提供が可能となる。

筆者らは、実験システムを試作し、この多重読取り原理の確認を済ませた[5]。最大アクセス時間46.7ms、読取り速度19.9Mbpsの磁気ディスク装置に、最大ビットレート2.05Mbpsの番組を蓄積した。端末での番組要求から表示開始までが0.8秒以下という高速応答の条件で、読取り多重度6の性能を得ている。

3.2 シングルディスクの多重度性能

1 タイムスロットで読取るセグメントのデータ

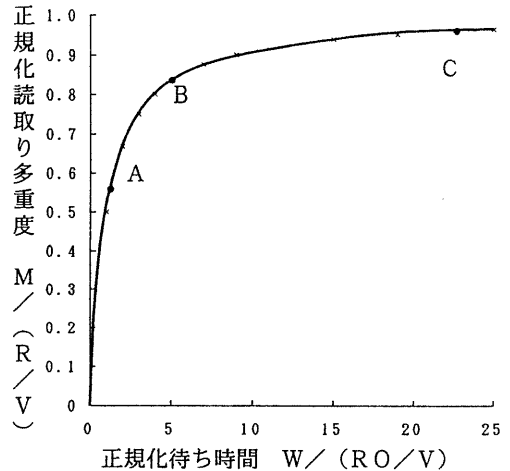


図4 シングルディスクの多重度

量をS、動画番組のビットレートをV、磁気ディスク装置の読取り速度をR、読取り処理のオーバーヘッドをOとすると、実現可能な読取り多重度の限界値Mは、

$$M = \frac{S/V}{S/R+O} \quad (1)$$

で与えられる。ただし、読取り処理オーバーヘッドOには、制御プログラムの実行時間、磁気ディスク装置のアクセス時間などが含まれる。

実現可能な読取り多重度は、厳密にはMを超えない整数であるが、以下では簡略化してMが整数となる場合について考える。M台の端末から同時に番組要求があり、最後に処理されることになった端末のために、最初のセグメントを読取るまでの時間を最大待ち時間と定義する。その値Wは、

$$W = S/V \quad (2)$$

で与えられる。式(2)を使って式(1)のSを消去して次式が得られる。

$$M = \frac{R}{V} \cdot \frac{VW/(RO)}{1 + VW/(RO)} \quad (3)$$

表1 磁気ディスク装置の性能例

項目	仕様
ディスク径	3.5 インチ
蓄積容量	4.1 GB
読取り速度	47~71.5 Mbps
シーク時間	0.6~1.7 ms
回転数	7200 rpm

式(3)を使って求めた多重度性能を図4に示す。点A、B、Cは、表1の磁気ディスク装置性能での設計例である。点Aは動画レート1.5Mbpsで待ち時間を1秒とした場合であり、多重度は17となる。点Bは、動画レートを6Mbpsに変更した場合であり、多重度は6に低下するが、R/Vでの正規化多重度は83%に向上する。すなわち、動画レートを4倍にしても、多重度は1/4までは低下しない。逆に点Bの設計点で動画レートを1/4にすれば、多重度、待ち時間共に4倍となる。

点Cは動画レート1.5Mbpsで待ち時間を18秒とした場合であり、多重度は30となる。カラオケのように番組再生中に次の番組を予約できるサービスでは、待ち時間を隠蔽できるので、多重度の点で有利である。また、図5に示すニアオンデマンドでも有利である。2時間の映画を3本蓄積し、各映画を12分ずつずらして10チャンネルで繰返し再生する。12分間に全番組データが磁気ヘッドのアクセスによってインタリーブして読取られる。これは、全番組の動画セグメントを予めインタリーブして蓄積しておき、内周から外周へ一方の読取りを繰返す文献[1]の方式と等価である。

シングルディスクの場合、待ち時間を増大することによって読取り多重度を大きくできるが、図4に示されるようにR/Vが限界値である。

3.3 並列読取り方式

並列読取り方式の構成を図6に示す。動画番組

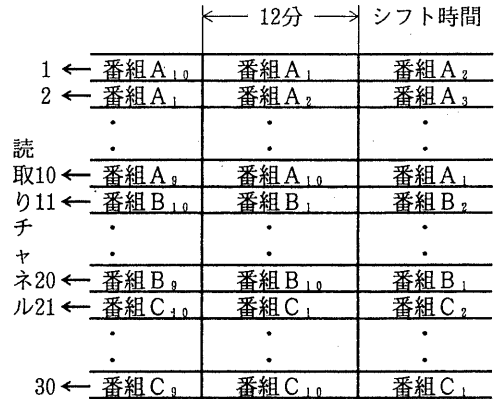


図5 ニアオンデマンド動作

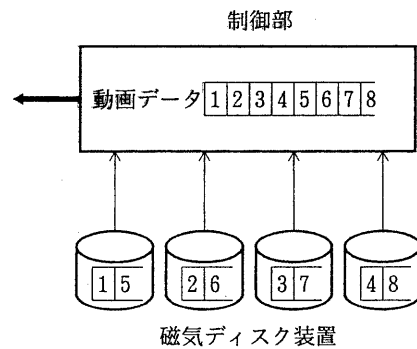


図6 並列読取り方式

を細切れに分割して複数の磁気ディスク装置に順繰りに蓄積する。読取り時には、全部の磁気ディスク装置から同時に読取りを行いつつ、データを元の順に並べ変える。これにより、読取り速度が磁気ディスク装置の台数倍に改善され、読取り多重度を向上できる。

筆者らは、蓄積時間が大きくて磁気ディスク装置台数が多くなる場合には高信頼化のためにRAID (Redundant Arrays of Inexpensive Disks)を利用し、逆の場合には低コスト化のために磁気ディスク装置を利用できる並列読取り装置を開発した[6]。SCSI (Small Computer System Interface)を4ポート有しており磁気ディスク装置4台の並列読取りができる。最大アクセス時間34.2ms、読取り速度25.5Mbpsの磁気ディスク装置を用いた8

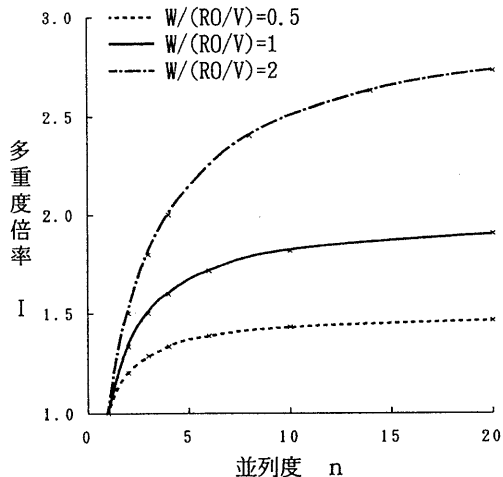


図7 並列度による多重度改善効果

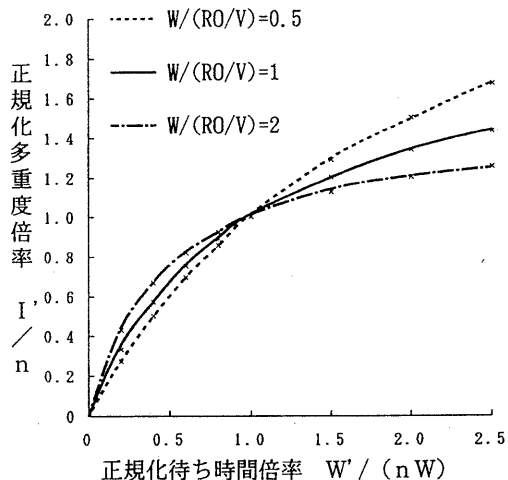


図8 待ち時間増大による多重度改善効果

多重の読取り実験で、56Mbpsの性能を得ている。

3.4 並列読取り方式の性能

磁気ディスク装置の並列度を n とすれば、読取り多重度 M' は、式(3)の R を nR に置換えて、

$$M' = \frac{W'}{VW' / (nR) + O} \quad (4)$$

で与えられる。シングルディスクの場合と同一待ち時間の条件で並列読取り方式にした場合の読取り多重度の倍率 I は、 $W' = W$ とおいて式(4)と式(3)の比で与えられる。

$$I = \frac{M'}{M} = \frac{1 + VW / (RO)}{1 + VW / (nRO)} \quad (5)$$

式(5)を使って求めた並列度による多重度改善効果を図7に示す。並列度の増大によって多重度が改善され、その倍率は待ち時間が大きいほど大きくなる事が分かる。しかしいずれの場合にも、多重度倍率は RO/V の正規化待ち時間に1を加えた値で飽和する。正規化待ち時間が2(表1の磁気ディスク装置の場合に実時間で1.6秒)の場合

には高々3倍留まりである。従って、さらに改善が求められる場合には、待ち時間を増大する必要がある。

待ち時間の増大を許容した場合、並列読取り方式のシングルディスクに対する多重度の改善倍率 I' は、式(4)と式(3)の比で与えられる。

$$I' = \frac{M'}{M} = n \cdot \frac{[1 + VW / (RO)]W' / (nW)}{1 + [W' / (nW)]VW / (RO)} \quad (6)$$

式(6)を使って求めた待ち時間の増大による多重度改善効果を図8に示す。待ち時間をシングルディスクの場合の n 倍にすれば、並列読取り方式は多重度を n 倍に改善できる。これよりも待ち時間を大きくする場合には、シングルディスクでの待ち時間が小さい場合ほど、実効読取り速度の改善が大きいため、多重度倍率が大きくなる。待ち時間を小さくする場合には、これと逆になる。

4. 高多重読取り方式

4.1 構成と動作

n ディスクの並列読取り方式は、シングルディスクと比較して、待ち時間を n 倍にすれば n 倍の

	← 大周期 →													
小周期	1			2				3			4			
タイムスロット	1	2	...	m	1	2	...	m	-----	1	2	...	m	
磁気ディスク 1	1	2	...	m	3m+1	3m+2	...	4m		m+1	m+2	...	2m	
磁気ディスク 2	m+1	m+2	...	2m	1	2	...	m		2m+1	2m+2	...	3m	
磁気ディスク 3	2m+1	2m+2	...	3m	m+1	m+2	...	2m		3m+1	3m+2	...	4m	
磁気ディスク 4	3m+1	3m+2	...	4m	2m+1	2m+2	...	3m		1	2	...	m	

図10 磁気ディスク装置と読取りチャンネルの接続関係 (n = 4 の場合)

多重度を得られることを節 3. 4 で明らかにした。この設計条件を利用した高多重読取り方式の構成を図 9 に示す。m 多重読み取り磁気ディスク装置 (または RAID) n 台が、スイッチを介して $n \times m$ 個の読取りチャンネルと接続される。読取り制御部は、磁気ディスク装置の多重読取りとスイッチ切替を制御する。符号化された各動画番組のデータは、先頭からセグメントに分割され、先頭のセグメントから磁気ディスク装置 1、2、...、n、1、2、... に順繰りに蓄積される。

n = 4 の場合を例に、磁気ディスク装置 1 ~ 4 と読取りチャンネル 1 ~ 4m との接続関係を図 10 に示す。タイムスロット 1 ~ m によって 1 つの小周期を構成して、磁気ディスク装置の数に対応した 4 個の小周期によって大周期を構成する。小周期 1 のタイムスロット i (i = 1 ~ m) で、磁気ディスク装置 j (j = 1 ~ 4) は、読取りチャンネル k (k = (j - 1) m + i) と接続される。以下、小周期が変わる毎に、磁気ディスク装置 j に接続される m 個の読取りチャンネル群が切替わる。読取りチャンネル 1 について見ると、小周期 1 では磁気ディスク装置 1、小周期 2 では磁気ディスク装置 2 というように、小周期毎に使用する磁気ディスク装置がサイクリックに切替わる。

本方式では、多重度 m の磁気ディスク装置 n 台で、読取り多重度を n 倍にできる。このとき、1 つの読取りチャンネルが 1 回で読取るデータ量は 1 セグメント分である。一方、節 3. 3 の並列読取り方式では、1 回で読取るデータ量はその n 倍となる。すなわち、本方式と並列読取り方式は、同じ

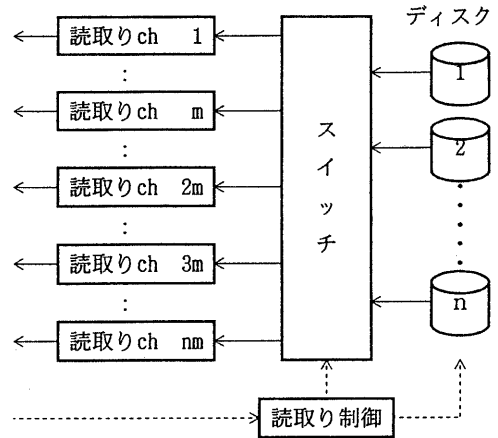


図 9 高多重読取り方式の構成

待ち時間の条件で同じ多重度性能を得ることができるが、バッファメモリ量の点で本方式の方が n 倍優れている。また、待ち時間の点でも、節 4. 3 で述べる改善策を採用することによって、n 倍優れているといえる。

4. 2 特殊再生

筆者らは、多重読取りにおける早送り等の特殊再生法を既に幾つか提案した [7]。しかし、高多重読取り方式では、使用する磁気ディスク装置がサイクリックに切替わるという特殊性がある。そこで、新たな特殊再生法、具体的にはスキップ再生とスロー再生を提案する。

図 11 に、スキップ再生とスロー再生の動作原理を示す。小周期は 1 秒と仮定する。スキップ再生

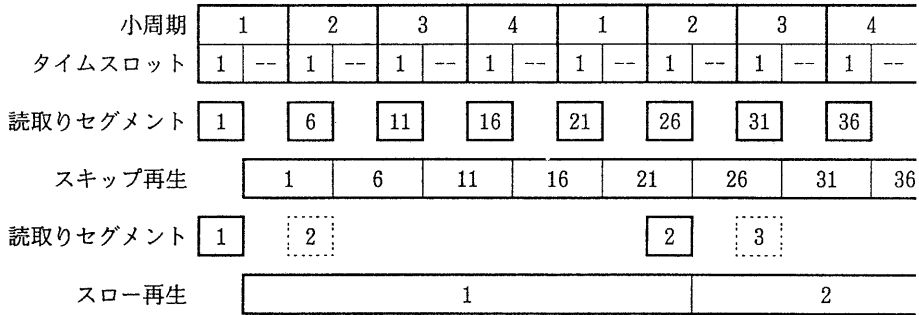


図11 スキップ再生とスロー再生

では、次の小周期で使用する磁気ディスク装置に蓄積されている後続セグメントを p 個スキップして読取る。端末では、1秒間通常再生しては $n \times p$ 秒間をスキップする動作が繰返される。これにより、 $n \times p + 1$ 倍の疑似早送りが実現できる。逆に過去のセグメントを $(p - 1)$ 個スキップして読取り、データを逆転再生すれば、 $n \times p - 1$ 倍の疑似逆早送りを実現できる。

スロー再生では、後続セグメントの読取りチャンスを p 回スキップし、かつ前セグメントのデータを $n \times p + 1$ 倍の時間をかけて再生する。逆に直前のセグメントの読取りチャンスを $(p - 1)$ 回スキップして読取り、かつ $n \times p - 1$ 倍の時間をかけて逆転再生すれば、逆スロー再生を実現できる。これにより、再生時間と読取りチャンスとの整合を図り、バッファデータ量の変動を少なくできる。

4.3 待ち時間の短縮

高多重読取り方式では、1小周期の時間内に、各読取りチャンネルは必ず磁気ディスク装置のいずれかと接続される。従って、番組要求時および特殊再生要求時に、下記の手順で処理することによって待ち時間を $1/n$ に短縮できる。

すなわち、本来アクセスすべき磁気ディスク装置の空きタイムスロットを小周期時間の範囲内で探し、空きが有ればこれを利用して読取りを実行する。もし空きが無ければ、最も近いセグメント

を蓄積している磁気ディスク装置の空きタイムスロットを小周期時間の範囲内で探す。以下同様に全部の磁気ディスク装置について探せば、小周期時間の範囲内で必ず空きタイムスロットが得られる。すなわち、本来要求されるセグメントから最大で $(n - 1)$ セグメント分のずれを許容することによって、最大待ち時間を小周期時間とすることができる。

5. 動画サーバ実験機

高多重読取り方式を検証するために、動画サーバ実験機を試作した。その概要を表2にまとめて示す。最大アクセス時間25.3ms、読取り速度34.3~56.5MbpsのSCSI磁気ディスク装置（3.5インチ2.1GB）を4台1組で使用した。スイッチには、VME (Versa Module Europe) バスを採用した。読取りチャンネルとしては、全体を1式のメモリボードと通信ボードで構成した。メモリボードから通信ボードへのデータ転送にはVSB (VME Subsystem Bus) を使用した。通信ボードは多重通信性能を飛躍的に改善したFDDI (Fiber Distributed Data Interface) ボード[8]である。

端末との通信はTCP/IP (Transmission Control Protocol/Internet Protocol) を採用している。そのフロー制御で動画伝送のレート制御を[9]、再送制御でエラー回復制御を行っている。図10における小周期は1秒に設定した。従って、待ち時間は1秒以下、動画番組の先頭欠落量は最悪でも

3秒である。特殊再生としては、順方向の可変速スキップ再生、スロー再生、コマ送り、およびジャンプと一時停止を提供している。1.5Mbps動画番組での実験で52多重の性能を確認した。

6. むすび

ビデオオンデマンドのセンタに用いる動画サーバでは、動画番組のコピーを必要としない条件での読取り多重度が重要な性能となる。読取り多重度について理論解析を行い、下記のことを明らかにした。シングルディスクの読取り多重度は、待ち時間を増大すれば改善されるが、磁気ディスク装置の読取り速度と動画番組のビットレートとの比が上限となる。並列読み取り方式による多重度の改善倍率は、待ち時間をシングルディスクでの設計条件数秒と同じにすると、高々数倍である。

さらに、並列度 n の条件で待ち時間を n 倍にすれば多重度を n 倍にできるという設計条件での高多重読取り方式とその特殊再生法、待ち時間短縮法を述べた。本方式は、並列読取り方式と比較してバッファメモリ量と待ち時間を $1/n$ に低減できる。本方式を採用した動画サーバ実験機を試作して、1.5Mbps動画番組の再生実験で52多重の性能を確認した。番組要求および特殊再生要求から提供開始までの待ち時間が1秒以下と、高速な応答を達成している。

謝辞

ご指導を頂いたメディア応用システム研究部の吉利誠部長に深謝します。

文献

- [1] W.D.Sincoskie : "Video on Demand: Is It Feasible?", Proc. IEEE GLOBECOM' 90, pp. 201-205(Dec. 1990)
- [2] 西村, 中野, 吉井 : "ビデオライブラリシステム", 特願平3-55310(Feb. 1991)

表2 動画サーバ実験機の諸元

項目	仕様
読取り多重度	52 (@1.5Mbps 動画)
待ち時間	1秒以下
通信方式	FDDI, TCP/IP
磁気ディスク	3.5インチ, 4台1組
特殊再生機能	可変速スキップ再生 スロー再生 コマ送り ジャンプ 一時停止

- [3] P.V.Rangan and H.M.Vin : "Designing File Systems for Digital Video and Audio", Proc. SOSP' 91, Oper. Sys. Rev., 25, 5, pp. 81-94(Oct. 1991)
- [4] H.Sakamoto, K.Nishimura, Y.Ishibashi and H.Nakano : "Multimedia Integrated Switching Architecture for Visual Information Retrieval Systems", Proc. SPIE Vol.1908, pp.123-132(Feb. 1993)
- [5] K.Nishimura, T.Mori, Y.Ishibashi and N.Sakurai : "System Architecture for Video-on-Demand Services", Proc. IEEE ICIP' 92, pp. 602-606(Sep. 1992)
- [6] T.Mori, K.Nishimura, H.Nakano and Y.Ishibashi : "Video-on-Demand System Using Optical Mass Storage System", Jpn. J. Appl. Phys., 32, Part 1, 11B, pp.5433-5438(Nov. 1993)
- [7] 石橋, 西村, 中野 : "ビデオオンデマンドサービスのための多重特殊再生技術の検討", 信学技報, CS92-74, pp.101-106(Dec. 1992)
- [8] 丸山, 中野, 西村 : "多重度の向上を目指したプロトコル処理技術の提案と評価", 信学技報, IN93-113, pp.45-52(Jan. 1994)
- [9] 西村, 森, 石橋 : "多重読取り可能なビデオオンデマンドシステム", テレビ学誌, 48, 3, pp.287-294(March 1994)