

大容量アーカイブを利用したコンテンツ流通の提案

丹野義和^{††††}, 石橋洋一^{††}, 伊藤学^{†††}, 露峰浩^{††††}, 青木輝勝^{†††††}, 新井健生[†]

[†]大阪大学, ^{††}山形県デジタルコンテンツ利用促進協議会, ^{†††}山形ナショナル電機㈱,

^{††††}松下電器産業㈱, ^{†††††}東京大学

概要

大容量映像の保存・管理は、大容量になるにつれて、オペレータ配置型が一般的である。最近のブロードバンド時代を迎え、アーカイブコンテンツを配信するに当り、コンテンツ ID による管理と、ユーザからのアクセスに対する配信能力を考慮した管理・検索方式を、保存棚と自動化ロボットとで構成した場合のコンテンツ配信能力を分析した結果から提案する。

Proposal of the Content Distribution Using Mass Archive System

Yoshikazu TANNO^{††††}, Yoichi ISHIBASI^{††}, Manabu ITO^{†††}, Hiroshi TSUYUMINE^{††††},
Terumasa AOKI^{†††} and Tatsuo ARAI[†]

[†]Osaka University, ^{††}Yamagata Digital Content Center, ^{†††}Yamagata National Electric Co., Ltd.,

^{††††}Matsushita Electric Industrial Co., Ltd., ^{†††††}Tokyo University

Abstract

As the volume of archives become larger, the archive system containing them is generally maintained and managed by operator. The distribution of archive content in the recent broadband era in mind, this paper proposes the management and retrieval system that realizes the management using contentID and the distribution dealing with the volume of accesses from users by preservation shelves and an automated robot. It also reports the result of analysis of content distribution capability of the system.

1. はじめに

近年の放送のデジタル化や、急速な IT の進展に伴い、新しいデジタル映像コンテンツの供給仕組み作りが大きな課題となっている。各放送局は、今までの映像資産の有効的活用方法を模索し、また新たに生み出される映像コンテンツも増加の一途をたどる。

[1]

また、各自治体でも、地域財産の保護や次世代への文化資産の継承などの動きから、資産をデジタル化し、アーカイブする動きが急激に進められている。更に、ブロードバンドや携帯端末の急速な進展に伴い、いちユーザが、情報の拡張手段として、映像などのコンテンツを配信できる環境が整ってきた。

このように、映像や静止面のコンテンツは増加の一途を辿っている。現実には、アメリカの 3 大ネットワークでの蓄積量は、メディア巻数にして ABC は 100 万本、NBC は 220 万本、CBS は 120 万本といわれている。また、アメリカのメジャーなスタジオでは、Disney 6,500 本、FOX 54,000 本、Sony/Columbia でも 54,000 本といわれている。日本国内においても、NHK アーカイブスでは、現在は 70~80 万巻ではあるが、近い将来 180 万巻~200 万巻がアーカイブされるとも

いわれている。

ここ、山形県デジタルコンテンツ利用促進協議会 (YDCC) でも、このデータセンターとなり得るための設備を有しており、7,200 巻 (SDTV 映像に換算して約 15,000 時間相当分) の容量を有するテープ自動装填ロボットを中心に、HDD サーバ、DVD サーバそれぞれを目的別に配置することによって、蓄積及び配信の効率を向上させるとともに、遠隔地からの検索、映像伝達、遠隔編集を可能としている [2]。

しかし、現在保有される映像資産は、アナログ VTR テープやデジタル VTR テープの形態で記録されているものがほとんどであり、これから生み出されるデジタル映像コンテンツは別として、この映像資産をデジタル映像配信に最適な形態に再編成させることは、莫大な労力と費用を必要とするため現実的ではない。実際、NHK が平成 15 年 2 月からにおいても、放送センターが保有する 70 万巻余 (全国の NHK では 160 万巻近くに達している) の映像テープは既存の映像資料として、放送に使われた媒体での保存が原則とされており、大規模なロボットによる倉庫等でなく、電動型収納棚に収納されたシステムと考えられる。 [3] [4]

従来からの収納棚は、基本的にあくまでも蓄積をしておくことが主目的であり、テープ資産を 2 次活用に

使いやすい仕組みとしては、ほとんど考えられていないのが現状である。

そこで、本論文では、映像資産は、オリジナルの映像テープのまま、データセンター的な施設の電動収納棚に保管され、ここに配信要求が集まることを仮定とし、蓄積を中心に考えられている収納棚の収納規模が拡大していった場合のランダムに抽出した VTR テープに対するアクセス能力を検証と有効な検索方式を評価し、大容量のアーカイブを利用した場合のコンテンツ流通を提案する：

2. 電動棚を使用したシミュレーション

2.1 基本ユニットの構成

ここでは、YDCC の設備である 7,200 巻自動装 ロボット[5]と、電動棚を組み合わせたシステムを想定し、保存メディアは DVCPRO フォーマットテープを用いたシミュレーションを行った。また、このシミュレーションでの保存媒体は、DVCPRO テープ使用し、1のはじめで示した主なメディアの最低必要保存巻数と、システムを構築した上での敷地面積を考慮し、テープ収納容量を 100 万巻以上と設定することにした。

この 100 万巻は、図 1 に示すような基本保存ユニットを 18,480 巻のテープ収納棚とし、それを 1 ブロックとして 64 ブロック複数配置していくことにより 1,182,720 巻のテープ保管棚を想定した。

このシミュレーション結果に、部分的に配信が集中するコンテンツに対して自動装填ロボットを適応した場合の配信能力の改善レベルと、実際に配信するに当たってのアクセス数への対応能力等にもおいても検証し考察を行った。

このテープ保管棚の基本ユニットには、DVCPRO テープが 42 巻収納可能な 1 列を横に 4 連結し、それを 10 段重ねて 1,680 巻収納できる棚を 11 棚つなげることで 18,480 巻収納可能なユニット構成とし、収納効率と運用性を考え、高い頻度のアクセスにも耐えられる電動開閉式とした。面積は縦 3,640mm、横 3,750mm の大きさになる。また、テープ 1 本あたり 1 時間としても、このスペースに 18,480 時間の映像を蓄積することができることになる。

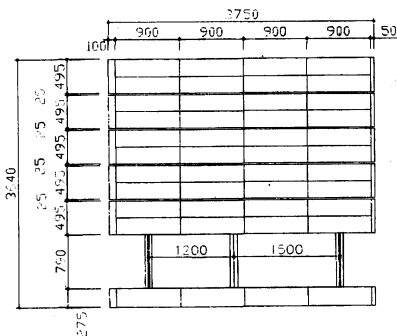


図 1 テープ保管棚の基本構造 (上面図)
Fig.1 Basic structure of the tape storage shelves

2.2 大容量テープ保管電動棚の構成

本シミュレーションにおける収納棚の拡張は、この基本ユニットを増加していくことで実施した。

拡張方法は、テープ再生装置の DVCPRO の VTR とモニターの設置部分を基点として、上記の基本ユニットを 1 ブロックとして縦、横それぞれに増加させた。配置は基本棚をべき乗で増加させる。棚と棚の間隔は人間の出入りを考慮して一定間隔を設ける。

各収納巻数における占有する面積を、表 1 各ユニットにおける棚面積に示す。また、配置レイアウトシミュレーション例として、図 2 に検証シミュレーション用 64 ブロックを示す。

表 1 各ユニットにおける棚面積
Table 1 Shelves area for each units

ユニット数	収納巻数	縦 (mm)	横 (mm)	面積
1	18,480	3,640	3,750	13,650,000
4	73,920	7,280	8,300	60,424,000
9	166,320	12,520	12,850	106,141,000
16	295,680	16,960	17,400	295,104,000
25	462,000	21,400	21,950	469,730,000
36	665,280	25,840	26,500	684,760,000
49	905,520	30,280	31,050	940,194,000
64	1,182,720	34,720	35,600	1,236,032,000

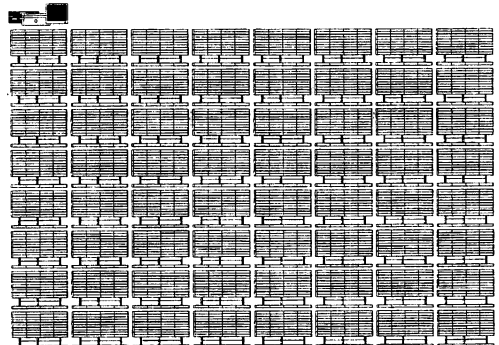


図 2 検証シミュレーション用 64 ブロック
Fig.2 64 blocks for validating simulation

3. 収納棚へのアクセス能力に対する検証方法および結果

3.1 基本ユニットにおけるテープの平均アクセス時間の検証

ここでは、基本ユニット内でテープをアクセスするために必要な時間を算出するための検証の定義を行う。図 3 のテープ保管棚アクセスシミュレーション図にて、アクセスするテープは 20 巻と定め、これが基本ユニット内の棚にランダムに存在すると仮定する。アクセスはオペレータが 1 名で行うものとし、基本ユニットの特定の 1 箇所 (本検証では、基本ユニット平面図左

上)を、基点と定め、この基本ユニット基点からスタートして所定の位置にあるテープを取り出し、基点に戻ってくるまでの時間を理論的に検証し、この検証をランダムに存在する 20 巻のテープに対して実施、この結果を平均して基本ユニット基点からの平均アクセス時間と定義した。

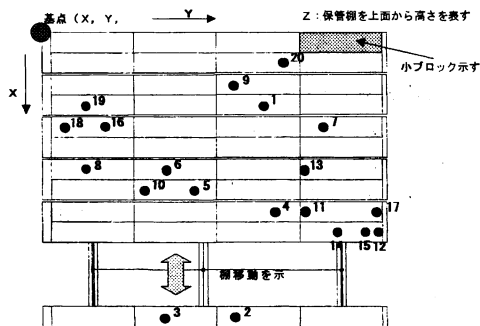


図3 テープ保存棚アクセスシミュレーション図
Fig.3 Access simulation of the tape storage shelves.

3.2 検証結果

図3の基点を原点として $X=0, Y=0, Z=0$ と設定し、20ヶ所のテープポイントを X, Y, Z 座標で表した。棚位置の X, Y の移動距離の単位はmmである。

基点から、各ポイントまでの往復距離を定め、人間の歩く速度を $1\text{sec}/\text{m}$ とした場合、各ポイントでの往復アクセス時間が決定する。この時間を平均した値が、表2の基本ユニットの平均アクセス時間は、約 7.8 秒となる。現実的には、アクセスポジションに到達してから、テープを視認、取り出す行動に対して、経験上 2 秒程度必要であり、この時間を加味した場合、基本ユニットでの平均往復アクセス時間は、10 秒と定めることができる。また、アクセスをポイント番号順に進めていった場合、電動棚の移動動作が発生する。この移動動作は、5 秒/回であるため、各ポイントのアクセスに対して、電動棚の移動が必要な場合は、この時間を往復アクセス時間に加え、各ポイントにおける総アクセス時間を求めることができる。

3.3 収納棚規模の変化に対する平均アクセス時間の検証

収納棚規模の拡張は、収納棚基本ユニットを1ブロックとして、縦、横に整数単位で増やす。従って、ブロック数は 1, 4, 9... という平方で増加させ、シミュレーション条件として以下を定義した。

- ・アクセス時間は、収納棚全体のスタートポイントから、各棚ブロックの基点に到達し、戻ってくる時の時間に基本ブロック内の平均アクセス時間を加えた値とする。
- ・全ての収納棚ブロックに対して 1 回ずつアクセスする。
- ・上記アクセス時間を平均して、収納棚全体における

平均アクセス時間と定義する。

- ・各棚ブロックに対して、1 度にアクセスする巻数を 1, 2, 3, 5, 10 巻とした場合についてそれぞれの平均アクセス時間を検証する。

3.4 検証結果

3. 3項で示した条件による 1 巻あたりの平均アクセス時間と保管巻数の関係を図4示す。

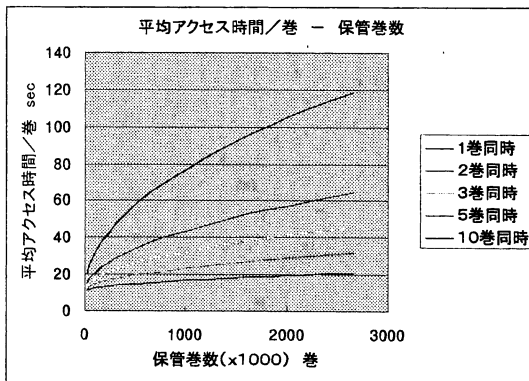


図4 1 巻あたりの平均アクセス時間と保管巻数

Fig.4 Access time average and number of tape storage per one tape

4. 検証結果に対する考察

4.1 シミュレーション結果に対する考察

図4から、100 万巻保管巻数の場合、1 巻だけの平均アクセス時間は、80 秒を必要とするが、同一ブロックで 10 巻まとめてアクセスするパターンでは、保管巻数増加の影響を大きく受けないことから、1 回で 10 巻以上のアクセスを行うようテープの配信をコントロールすることにより、本モデルにおいて、15 秒程度のアクセス時間でテープの取り出しができることになる。実際の運用においては、これにテープのケース脱着等の雑時間 (平均約 15 秒、これは習熟度合いによって 10~20 秒の間変動する) が加わるため、習熟度が最高に達した作業者にとって約 30 秒/巻の平均アクセス時間が必要と考えられることがわかった。

4.2 実際の映像配信要求分析

テープ保管を行う映像データセンターを仮定した場合、今後の映像コンテンツの増加およびブロードバンドを含む高速ネットワークの普及、これに著作権に対するの保護の確立によって、このデータセンターの配信要求は急速に高まり、MPEG-7 標準記述を用いた検索技術等がデータセンターにあるコンテンツ検索の大きな助けとなることが、容易に想定できる為、配信要求は今後大幅に増大することが予想される。[7]

また、映像配信要求は、保存コンテンツにすべてに発生するわけではなく、その時の人気やトレンド

ドにより、大きな偏りが発生することが考えられる。この配信要求分布は、インターネットプロバイダーに対するキーワード検索分布と同様な傾向をもつと考えることができる。データを分析した結果、ジャンルにかかわらずほとんどのサンプルが $R^2=0.936$ と極めて精度よく累乗関数に近似できることがわかった。これを基に、蓄積数 100 万巻のコンテンツに対するアクセス数を求めた。

アクセス順位 1 位のコンテンツの指数を 100 として、2 位以下を相対的に表わすことより、1 位のコンテンツの指数を 1 として、相対評価を考えた累乗関数は、 $y_n = T x^{-0.55}$ の基礎式で表される。次に、 $x=1$ の時、すなわち最大アクセス数を持つコンテンツのアクセス回数を T と定め、任意の順位に位置するコンテンツのアクセス数 Y は、 $Y = T x^{-0.55}$ と表わすことができる。

一方、蓄積コンテンツ数 100 万巻に対して 1 日あたり 28,000 件のアクセスがあることを仮定すると、 $x=1, x=1,000,000$ で定積分した値が 28,000 に等しくなる。したがって本検証モデルにおいて、 n 番目の順位のコンテンツに対する 1 日のアクセス回数は、 $Y=25.2 x^{-0.55}$ と表す事ができる。

4.3 YDCC 所有アーカイブ装置との組み合わせの分析

7,200巻のアクセスに関して、倉庫部7,200巻、送出部100巻とし、検索上位順位が全て送出カートにあると仮定する。上位100巻まで交換回数は、この式で、順位 1 から 100 位まで積分を行うと、 $25.2/0.45(7.9-1)=388.8$ 回 (移送時間 10 秒) になる。ロボット全体 (7300 巻) での交換回数は、同様にして順位 1 から 7300 まで積分を行うと、 $25.2/0.45(54.8-1)=3010.8$ 回となる。従って、倉庫部からの交換回数は $3010.8-388.8=2622$ 回 (実測平均値 70 秒) となる。全体の平均能力として表せる倉庫のアクセス時間を求めると (倉庫部 70 秒、送出部 10 秒として) $(388.8 \times 10 + 2622 \times 70) / 3010.8 = 62.3$ 秒となる。

倉庫・送出 7300 巻の中で、100 巻 (1.4%) のテープを送出カートにおいて運用することにより、総アクセス数=3010.8 回のうち 388.8 (13%) について高速アクセスを提供できることがわかる。

100 万巻における総アクセス量が 28,000 回/1 日として、自動ロボット 7300 を使用した場合、電動棚は、992,700 巻を受け持つ。7300 番目以下のアクセス総数は、 $x=1-7300$ 巻の定積分で求め、3010.8 回となる。このことから、100 万巻の電動棚に対して 1 日に発生するアクセス量は 28000 回 - 3010.8 回 = 約 25000 回となる。24 時間で処理する場合、1 コンテンツあたり 3.46 秒が必要となる。このことから、電動棚に対する一人で行う 10 巻あたりのアクセス時間は 18 秒 (予想値) $18/3.46=5.2$ つまり約 6 人が必要である。

以上のことより、この保管棚における配信要求対応に対して、ほぼ短期間に発生する繰り返し配信のコンテンツについては、自動化管理を行うことにより、非常に大きな効果が得られることがわかる。たとえば例として、上記 1 週間での繰り返しが発生するコンテ

ンツ数が 5,000 件と考えると、この 5,000 件分のバッファをもつ、アーカイブ自動配信システムが存在することで、100 万巻以上の映像保管システムが非常に有効に活用されることになるかと考えることができる。実際には、配信運用統計から自動配信システムの規模と、そこでの保管期間を定めることになるが、この自動配信システムの規模は、総配信要求量に依存し、この総配信要求量は、映像資産の蓄積規模に依存するという関係が存在すると考えられる。

5. cidF 準拠 cid 用いたコンテンツ流通

5.1 cid の導入とシステム構築

大容量の映像コンテンツを効率良く管理し、健全に流通させるには、効率検索や著作権保護等の配信環境整備が重要な課題である。そこで cidF (コンテンツ ID フォーマット) [6] の提唱する cid 付与方式を適用した ID 管理センタと透かしセンタを取り入れた大容量型映像アーカイブ管理システムを再構築し、機能評価をした。

評価内容は、今後の映像コンテンツ流通において、盛んに使用される MPEG-4 圧縮動画 (ここでは、Windows Media) を用いたコンテンツに、電子透かしによって、コンテンツ ID とシーン毎にオリジナルタイムコードを埋め込み、抽出を行い、その抽出による時間測定、画質評価を行った。図 6 に山形オリジナルで、かつ cidF 準拠のコンテンツ ID を示す。この方法により、映像の各フレームが世界で唯一のユニークなコードとなり、オリジナルタイムコードの抽出により、遠隔地での編集をも可能とした。この ID の特徴は、

- ①素材・完パケそれぞれ 20 億のタイトルのコンテンツと、あわせて映像の各フレーム単位の管理と流通が安心して行える。
- ②映像素材と完パケ (番組) の区別ができる。 [7] [8]

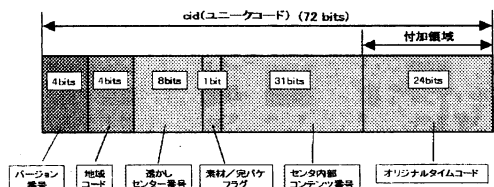


図 5 山形オリジナルのコンテンツ ID 付与方式
Fig.5 Access time average and number of tape storage per one tape

5.2 透かしコンテンツの仕様

アーカイブ収納素材のベースバンド映像を原画コンテンツとして使用し、ID 管理センタより付与された ID センタ管理番号を、固有の電子透かし方式にて配布されたコンテンツに埋め込み MPEG-4 フォーマット信号を生成した。ここでは、MPEG-4 のコーディングは、I-VOP と P-VOP のみで符号化を行った。画像サイズについては、320×240 の CIF サイズに対して 7fps で MPEG-4 の符号化を行い、ビットレートは 768kbps としている。

5.3 実透かし検出サーバの構成

透かし検出サーバの構成は、DCA j 様が先に行った「コンテンツ ID の共同検証実験」に参加した際に使用したのと同じ構成にし、独自に MPEG-4 対応に改定させた実透かし検出サーバを構築し、クライアントソフトウェアについては簡易的な端末ソフトウェアで対応している。

ユーザ端末の GUI 上で選択された画像のキーフレームの区切りを単位として切り出された画像に対して、MPEG-4 対応透かしサーバ内の抽出システムで実透かしの抽出を行う。MPEG-4 の切り出しは、ツールを用い、キーフレーム単位で MPEG-4 ストリームを切り出し、これを実透かしサーバで抽出を行った。

実透かし検出サーバ内で、切り出されたキーフレーム単位での MPEG-4 ストリームから実透かしの抽出が終了すると、端末上には抽出結果が表示される。

この GUI 上には、抽出されたコンテンツ ID とオリジナルタイムコードが表示できる。今回の端末ソフトウェアでは、抽出結果を表示し、抽出されたコンテンツ ID に該当する映像のメタデータを取得し、同じく利用者端末上に自動表示するしくみである。[9]

6. MPEG-7 と cid の融合

YDCC では、デジタルコンテンツ検索の整備構築として、一般性のある検索記述手法の導入を目的とした MPEG-7 記述方式を取り入れた。ここでは、検索記述内容を、MPEG-7 Part.5 MDS [10][11]のスキームに当てはめ、1 つのシーンに含まれるオブジェクト毎の名詞句を記述することにし、検索語付与の最小単位は“object”とし、オブジェクトをまとめた単位を AudioVisual id=“Scene-n (n=1.2.3....)”とした。更に、シーンをまとめた単位を AudioVisual id=“program”とした。

これらを、映像素材全体に共通な記述として、cid、タイトル、素材分類、キーワードなどをまとめ、シーンにはオブジェクト記述を行い、階層構造にしオリジナルの MPEG-7 準拠検索プロファイルの構築と、Web 上で検索するための実験用検索アプリケーションを開発した。代表画面を図 6 に示す。[12]

また、デジタルコンテンツ制作・流通環境の整備構築を目的とした cIdf 準拠の属性情報のそれぞれのデータを一緒にせず別々管理し、必要な時にリンクさせる方法を用いた。両方式を有機的に結合するためには、両 DB 共 XML 構成されていることを利用し、cIdf で用いている cid を共通キーとすることとした。このことで、cid を通して相互乗り入れが可能になり、ユーザにとって双方からのデータアクセスが容易となる[13]。

7. まとめと今後の課題

以上の検証、考察より、以下の点が明らかになった。電動棚の規模が大きくなると、1 巻のみに対してのアクセス時間は、そのまま増加するが、本論文で述べた

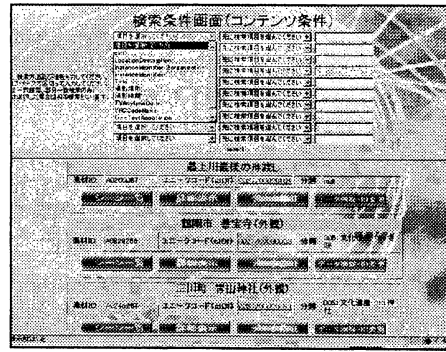


図 6 MPEG-7 準拠検索ブラウザ
Fig.6 GUI of MPEG-7-based retrieval application

テープアクセス方法を工夫することによって、テープ保管規模のアクセス時間に対する影響を、非常に少なくすることができる。

- ・100 万巻電動棚での平均アクセス時間は、テープアクセス方法を工夫したとしても平均 30 秒必要とする。
- ・配信に対しては、ある期間での繰り返し配信量以上のコンテンツを定め、これを自動配信とすることで、100 万巻以上の映像保管に関しても、効率的な配信を行うことができる。
- ・アーカイブ自動配信の規模は、総配信要求量に依存して決定され、総配信要求量は、蓄積規模に影響を受けると考えられる。
- ・cIdf 準拠 cid 付与方式を導入することによって、健全な流通が可能になった。
- ・MPEG-7 と cid を融合管理することで、検索・管理の向上がはかられた。

ここ YDCC には、映像素材が 3,000 タイトル存在する。ビジネスとして流通されるコンテンツも教育コンテンツ・観光コンテンツ合わせて 300 タイトルが登録された。今後、更に増えつつ流通される番組が安全でかつ効率よく管理される為に、HDD とのハイブリッドなアーカイブを検討していく。

【参考文献】

- [1] 総務省編：“情報通信白書 平成 13 年度版，” pp.212-222, 266-267, 2001.
- [2] 長谷川文雄，前原文雄，川畑 優，比企春夫，丹野義和，伊藤 学：“大容量アーカイブ活用型番組制作システム，” 信学技報，IIE99-83，Nov.1999.
- [3] 沼澤潤二，上原 年博：“デジタルマルチメディア時代の映像アーカイブ技術”，NHK 技研だより，P7-14, Mar.1999
- [4] 加藤 隆：“映像アーカイブ用動画検索システム～大規模映像資料の高速検索を目指して～” 平成 11 年 技研公開 講演・研究発表 予稿集，P43-48, May 1999
- [5] 丹野義和，前原文雄，関谷里美，伊藤学，露峰浩，長谷川文雄“大規模自動装填ロボットにおけるペタバイト級アーカイブシステム”：信学論，D- II Vol. J84-D- II No.6 pp.1102-1111 (2001)

- [6] コンテンツ ID フォーラム. (cIDf) : www.cIDf.org/
- [7] 比企春夫, 伊藤学, 長谷川文雄 “MPEG-7 記述によるアーカイブ映像活用方式” 映情学誌, 57,2, pp-272-276 (Feb.2003)
- [8] 安田浩/安原隆一監修 “ポイント図解式コンテンツ流通教科書” 株式会社アスキー出版
- [9] 川畑優, 長谷川文雄 “電子透かしによる画像劣化の一検討 (2)” : 映情学年次大会, 11-2, Aug. 2002
- [10] S. Manjunath, P. Salembier and T. Sikora : “Introduction to MPEG-7 - Multimedia Content Description Interface” , John Wiley and Sons. (2002)
- [11] ISO/IEC FDIS 15938-5 : “Multimedia Content Description Interface - Part 5: Multimedia Description Schemes” , JTC1/SC29/WG11/ N4242, (Oct. 2001)
- [12] 伊藤学, 比企春夫, 小暮拓世, 石橋洋一, 長谷川文雄 “MPEG-7 を用いた素材映像検索アプリケーションの検討” 映情学誌, 57,7, pp-812-818 (July.2003)
- [13] Yoichi Ishibashi, Takuyo Kogure, Yoshikazu Tanno, Manabu Ito and Fumio Hasegawa: “Verification Tests Approach based on “MPEG-7 over cID” for the Large Scale Video Archive - IPR Management and Subjects of Contents Distribution - “ IWAIT2003, pp236-239, 21-22, Jan 2003