

動画像データベースの格納構造

清光 英成 山本 憲男 高倉 弘喜 植村 俊亮

奈良先端科学技術大学院大学 情報科学研究科

高速ネットワーク技術，計算機技術の発達によりマルチメディア情報の生成と利用が活発に行われている。特に音声と動画像は人間が最も直感的に事象を把握できるメディアであり，データベースシステムにおいてもその応用技術が盛んに研究されている。本稿では，オブジェクト指向データベース管理システムを用いて動画像をデータベースにオブジェクトとして格納し，ユーザの問合せにより所望の場面映像を検索・合成・再利用できる動画像データベースシステムの実現を目的とする。そのため動画像データを意味構造，論理構造，物理構造に分けてモデル化し，データベースに蓄積する格納構造を考察する。

A Structure of Video on Database Systems

Hidehiko KIYOMITSU, Norio YAMAMOTO, Hiroki TAKAKURA and Shunsuke UEMURA

Graduate School of Information Science
Nara Institute of Science and Technology (NAIST)

We propose a structure of Video data on database systems. In this paper, the structure is recognized by three aspects of the data conceptually, semantic, logical and physical. The first consists of meanings of the time intervals. The second consists of such conceptual constituents as "act" or "scene" a movie. The last consists of physical constituents, for instance "stream", "pack", "packet" in MPEG. In the storage structure, we modify the streams of MPEG in order enable to individual synchronized playback operation. And we discuss an algorithm to implement the playback operation with synchronization.

1. はじめに

CAI や医療情報処理, 事務情報処理など多くの応用分野で図版や動画などのマルチメディア情報を扱う要求が高まっている。今日, ATM や Fast Ethernet などの高速な通信を可能とする技術に加えてディスクアレイに代表される大容量の記憶媒体の開発により, マルチメディア情報システムの研究・開発が盛んに行われている。また, データベース研究の長年の成果からマルチメディアデータベースシステムの実現が可能となりつつある [増永 96a] [増永 96b]。

テキストや数値データと比較して動画像メディアのデータは, はるかに大容量である。この問題に対して, MPEG をはじめとする動画像圧縮符号化アルゴリズムを用いて記憶領域やネットワーク負荷を削減する手法がとられている。しかしながら, 動画像データをデータベースに格納し, 動画像の内容や構造に対する多様な検索をサポートするためには, 論理構造と意味構造のモデル化が必要である。動画像中の登場人物・動作等は二次情報の付加によってデータベース上に表現されており, 論理構造を構成する要素間をまたいでいることがある。この場合, 構造化文書などの論理構造で議論されていないデータ定義の問題がある。また, 動画像メディアデータは音声メディアやテキストメディアのデータと時間的・意味的に密接な関連性を持つことで, 人間にとって最も把握しやすい表現を提供している。そこで, 本稿では動画像データの構造のモデル化と格納構造を議論する。またそこでは, 圧縮符号化アルゴリズム MPEG-1 を用いて符号化された動画像データをデータベースシステムに格納した場合の同期再生アルゴリズムについて議論する。

2. 動画像データの構造

2.1 論理構造と意味構造

演劇は大きくは, 幕・場から構成される。そして, 幕は一つ以上の場で構成されるという部品展開構造を持ち, その中で第1幕の次は第2幕であり, 第5幕は第1幕よりも後であるという順序関連をもつ。演劇などを撮影して動画像オブジェクトを生成する場合には, 幕は場を表現する動画像オブジェクトを順序関連をもとにつなぎ合わせた複合オブジェクトであり, 演劇を表現する動画像オブジェクトは幕を表現する動画像オブジェクトを順序関連をもとにつなぎ合わせた複合オブジェクトであると捉えることができる。これは幕と場の間にもあてはまる。この構造を本稿では, 論理構造と呼ぶ。動画像データの

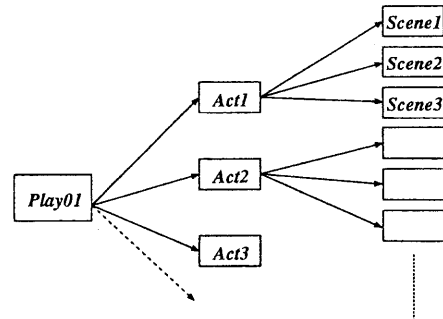


図1 演劇の論理構造

内容・意味は登場人物や大道具・小道具などの物体とその動きや発声を場面として捉えたものであり, 構造化するためには合成演算などの前処理を必要とする。場面は格納構造における動画像データの該当する閉時区間 [All83] を写像した意味構造の構成要素であり, 任意のラベルを付けることで検索の単位としてデータベースに格納される。動画像の意味構造では場面が動画像データの論理的な切れ目や物理的な切れ目をまたぐ事象がしばしば見られる。

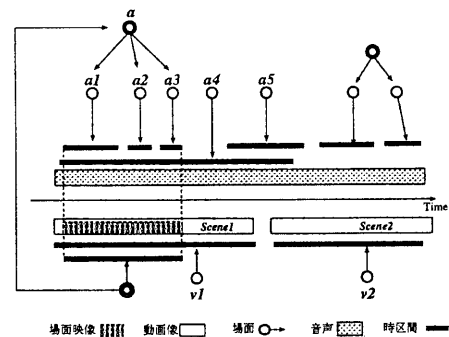


図2 音声と動画像の意味的な関係

例えば図2で a_5 と *Scene1*, *Scene2* の関連がこれにあたる。逆にこの a_5 の持続時間内で動画像が信号的に「空」となるとも捉えられる [福田 94]。このように, ある事象を表現することはその事象を捉える側面に大きく依存している。したがって, 音声付きの動画像データでは動画像メディアのデータと音声メディアのデータは個別の構造を持ち, 互いに関連し合っていると捉えるのが自然である。また, 音声

と動画は索引づけされた時区間を同期的に共有する。音声に対する索引から会話場面などを合成しようとする、図2のaのようなになる。aは $during(a, a1) \wedge during(a, a2) \wedge during(a, a3)$ を満たす閉時区間で表現され、また、動画データ中で対応する場面を特定する。ここでは、時区間という側面で動画と音声を統合できたことになる。本稿では、このような場面と複数の場面の合成から得られる意味のある場面を意味単位と呼び、意味単位と合成過程を総称して意味構造と呼ぶ。

動画データの内容を検索で扱うための研究は多くなされており、被写体に依存した意味的な区切りを論理カットという閉時区間として扱う手法や、被写体の動きに追従するために時区間と画像の対象領域をメタ情報として保有するホットスポットなどの手法が提案されている [有澤 93][HKTH94]。これらの研究は動画データの内容構造に基づく多様な検索機能を実現できることが期待されている。

2.2 物理構造

計算機上で動画は、1秒間に連続する30枚の静止画像を表示することによって表現される。1画素を24bitで量子化して 320×240 画素の大きさの動画を表現しようとする、1秒あたり6.75MBの記憶領域を必要とする。動画データを記憶装置に格納したり、ネットワークを用いて送受信する場合は容量の問題を考慮して圧縮符号化を施すことが一般的である。本稿では、動画の圧縮符号化アルゴリズムであるMPEG-1を用いて動画情報情報を格納することを前提にして物理構造を捉える。

タイプ	I	B	B	P	B	B	P	B	B	I	B	B	P	B	B	P
入力順	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
出力順	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
タイプ	I	P	B	B	P	B	B	I	B	B	P	B	B	P	B	B
表示順	1	4	2	3	7	5	6	10	9	8	13	11	12	16	14	15

図3 フレームの入力に対するMPEG符号器からの出力順と対応するピクチャタイプの例

MPEG-1 動画は基本構成要素として連続するピクチャ(picture)と、異なるタイプの複数のピクチャがある規則に基づいて整理させたGOP(Group of Pictures)をもつ。MPEG-1 動画のシステムストリームはパック(Pack)の列からなる。パックはパックヘッダ、パケット(Packet)の列からなり、ストリームの最初のパックはシステムヘッダを持ち、

最後のパックはエンドコードで終了する。動画1フレームに相当するAU(Access Unit)は複数のパケットに渡って表現されている。

3. 動画データのモデル化

前述のとおり、動画データは部品展開を抽象的に捉えた論理構造と物体や動作などを概念的に捉えた意味構造を持っている。本稿では特に論理構造について議論する。

3.1 論理構造のモデル化

動画データをデータベースシステムで扱うために、前節で述べた論理構造を記述できるデータモデルを提案する。部品展開の構造を明示的に記述するために、複合オブジェクトに包含される被参照オブジェクトの定義を一般の属性定義と分けていことに注意したい。本稿ではオブジェクトを基本オブジェクトと複合オブジェクトの2種類に分けてい。基本オブジェクトは実際にデータの内容を持ち、複合オブジェクトは基本オブジェクト又は複合オブジェクトを構成要素として一つ以上持つ。

3.2 動画オブジェクトの定義

動画メディアは時間と共にその属性値を変化させる性質を持つので、動画メディアデータで構成されるオブジェクトは、開始時刻と終了時刻の時刻属性を持つ時区間オブジェクトとして定義される。例として演劇をモデル化する。

```
class Play {
  Attribute :
    AbsoluteTime stp, etp;
    string titles;
    Play_Audio audio;
    Play_Video video;
  Operation :
    playback();
  .....
};
```

演劇を表現するクラス Play は属性として題名、開始時刻と終了時刻を持ち、音声と動画という二つのメディアから構成されるので、演劇を音声の側面で捉えたクラス Play_Audio と 動画の側面で捉えたクラス Play_Video 型の変数を属性に持つ。

```

class Play_Video {
  Attribute :
    AbsoluteTime stp, etp;
    string titles;
    Act acts[ ]
  Operation :
    playback();
    setAct(Act);
    .....
};

```

演劇の動画の側面を表現するクラス Play_Video は複合オブジェクトクラスである。属性として開始時刻と終了時刻を持ち、幕を表現するクラス Act の可変長配列を持つ。Act 間の順序関連はクラス Act で定義され、一つ以上の幕からなるという制約はクラス Play_Video の格納時に検査されるように定義されている。動画データの実体はクラス Act のオブジェクトを構成する Scene オブジェクトが持ち、Scene オブジェクト間の関連から Act オブジェクトが生成できる [増永 96b]。

4. 格納構造

動画データをオブジェクト指向データベース管理システム (OODBMS) を用いてデータベースに格納する。MPEG-1 で圧縮符号化された動画データをオブジェクト化する時の問題点は

- 1) 符号化前のフレームの順序と符号化後のフレームの順序が異なる。
- 2) I ピクチャ以外のピクチャはフレーム間相関を用いた予測符号化を行っているため単独で復号できない。また、フレーム間相関を用いた予測符号化を行っているピクチャの復号が直前の GOP に依存する場合がある。

などであった。これらの問題は、MPEG-1 のシステムストリームに編集を加えることで解決できる [小川 96]。本稿では、部品展開構造を表現できる定義言語を導入すると共に、MPEG-1 のシステムストリームを真にオブジェクトとしてデータベースに格納する。さらに格納された動画オブジェクトの再生時の同期を議論する。

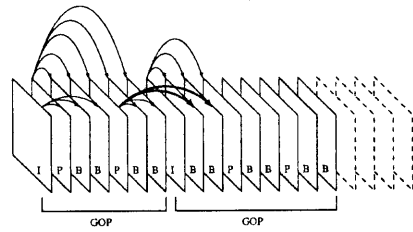


図 4 MPEG-1 圧縮符号化動画の格納時のフレーム順と動き補償予測

4.1 MPEG-1 システムストリームのオブジェクト化

MPEG-1 圧縮符号化動画データへのランダムアクセスを可能にするシステムストリームに対する編集をストリームの生成時に行って、システムストリームをオブジェクト化する。以下にシステムストリーム編集の主要部分を述べる。

- i) 動画 1 フレームに相当する AU (access unit) の開始点はバケットヘッダの直後、AU の終了点はバケットヘッダの直前となるようにバケットを再構成する。AU の開始点の直前に挿入されたバケットヘッダのパラメータにタイムスタンプ情報を付加する。
- ii) I ピクチャの高速再生のため I ピクチャの開始点はバックの開始点、I ピクチャの終了点はバックの終了点にそれぞれ一致させる。また、I ピクチャの開始点の直前に挿入されたバックヘッダにタイムスタンプ情報を付加する。
- iii) システムストリームを GOP 単位に区切る。また、システムストリームヘッダを単独の部分ストリームとする。

格納する動画を GOP に対応するストリームの順序集合ととらえてオブジェクト化する。動画全部を表現するシステムストリームをクラス SysStream のオブジェクトとすると、クラス SysStream は属性にシステムストリームヘッダを格納しているバックとクラス GAU 型の可変長配列を持つ。クラス GAU のオブジェクトは該当する GOP のストリームを表現するビット列を属性値として持つ。

```

class SysStream {
  Attribute :
    InternalTime duration;
    Pack stream_header;
    GAU stream[ ];
  Operation :
    playback(InternalTime tp);
    .....
};

```

システムストリームヘッダを格納しているバックとGAUを分けて格納することにより、データベースに格納された動画像データにランダムアクセスすることができる。また、取り出した任意の動画像部分の先頭にシステムストリームヘッダを格納しているバックを付加することにより、格納されている動画像データの途中からの再生も実現できる。

4.2 MPEG-1 動画像の再生

動画像データベースシステムでは動画像オブジェクトが2次記憶装置に格納されているのが普通であり、これらの記憶装置から一定時間内に読み出すことができるフレーム数は計算機、ネットワーク、周辺機器の負荷に依存する。動画像オブジェクトは再生が開始されて t 秒後には $30t+1$ 番目のフレームを表示しなければならない。そこで、動画像オブジェクトの個体同期再生操作関数を定義する為に時間待ちとコマ飛ばしのアルゴリズムを以下に示す。 f_i を第 i フレーム (i は自然数)、 g_n を第 n GOP、 $g_{n.e}$ を第 n GOP の終了予定内部時刻、 $g_{n.j.num}$ を第 n GOP 中の第 j 番目のフレームの順番、 dt を伸長にかかる時間 (sec) (初期値 0)、 $current_time()$ を現内部時刻 (sec) を返す関数とする。

f_1 を復号する。 i に 2 を代入する。 $current_time()$ が内部時刻を評価するための基準時刻を設定する。

f_i を表示する。ストリームの最後まで (A) を実行する。

(A) f_i を復号する。

a) $current_time() - dt \geq 1/30$ ならば、 $g_{n-2.e} < current_time() + 2dt < g_{n-1.e}$ が真となるように n を設定する。

i) g_n のストリーム編集フラグに closed_gop フラグが立っていないならば broken_link フラグを立てる。

i を $g_{n1.num}$ に設定する。 f_i を復号して表示する。

b) $i - 1 \leq current_time() \times 30 < i$ が真ならば、 f_i を表示する。

c) $current_time() \times 30 < i - 1$ が真ならば、 $current_time() \times 30 \geq i - 1$ が真になるまで待つ、 f_i を表示する。

i に 1 加算する。 dt に $current_time()$ が返す値を代入する。

このアルゴリズムの特徴は、表示が遅れる場合はフレームの伸長を順次実行し、コマ落しを行っている。また、伸長が $1/30$ 秒よりも長引いた場合はフレーム単位のコマ落しでは結果的に 1 フレームも表示できないことが推測できるため、次の GOP ストリームから伸長を再開するというものである。このアルゴリズムでは、格納時のシステムストリームの編集が有効に機能しており、GOP 内で予測符号化が可能な closed GOP とそれ以外の GOP の両方で個体同期再生が可能となる。

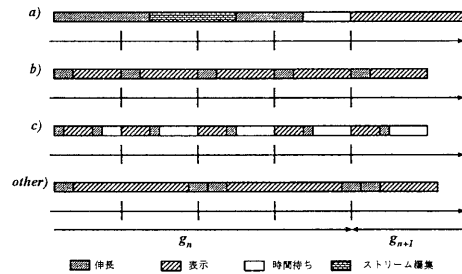


図 5 同期再生機構の概念

5. まとめ

動画像データベースシステムをオブジェクト指向データベース管理システムを用いて生成するために、本稿では特に論理構造、物理構造と格納構造について述べた。動画像データの論理構造は意味構造と密接な関連があり、構造化文書データベースの研究で議論されていない問題が今後明らかになると考えられる [Ae94]。動画像データはその容量が非常に大きいため、圧縮符号化アルゴリズムの一つである MPEG-1 を用いて符号化する。ここで問題となるのは、符号化された MPEG-1 システムストリームにランダムアクセスし、再生時の同期を保証するためには、任意の GOP から再生を始める機構と復号処理において遅延を生じた場合、複数のピクチャをコマ飛ばしできる機構が必要であるということである。そこで、

MPEG-1 システムストリームを GOP 単位でオブジェクト化する前にシステムストリームに対して編集を加えることにした。論理構造の記述法についてはモデルのレベルで構成要素間の論理的な制約を記述できる言語が必要となるが、今後の課題として、別個に記述される他のメディアのオブジェクトとの関連記述や、複数種類の構成要素クラスを複合要素として記述するための問題を解決する予定である。また、分散環境に対応できる機構にも拡張する予定である [端山 96]。動画像データの意味構造は、その意味を持つ時区間に対して自動的に索引を付けることが困難であるため特に触れていないが、有益な成果を上げている研究が多く見られる [八幡 95]。メディアの異なる複数の時間的オブジェクトの同期再生は大きく分けて並列再生、直列再生がある。HyperODA や MHEG では parallel, sequential と呼ばれ、空時区間 delay を用いることでマルチメディア同期を実現できるといわれる [BE95]。マルチメディアデータベースシステムの実現を目指し、これらの標準化にも符合するモデルおよび実装とするつもりである。

<謝辞>

日頃から御議論・御指導頂いている植村研究室の皆様へ感謝します。本研究に精粋の御示唆を頂いた図書館情報大学教授増永良文先生に心から感謝します。

参考文献

- [Ae94] W. Appelt and (ed.): *2nd DIS ISO/IEC 8613-14/Draft Recommendation T.424, Information technology - Office Document Architecture(ODA) and interchange format - Temporal relationships and non-linear structures*, 1994.
- [All83] J. F. Allen: "Maintaining Knowledge about Temporal Intervals", *Comm. of the ACM*, Vol. 26, No. 11, pp. 832-843, November 1983.
- [BE95] T. M. Boudnik and W. Effelsberg: "MHEG Explained", *IEEE MultiMedia*, Vol. 2, No. 1 spring, pp. 26-38, 1995.
- [HKTH94] 平田 恭二, 川崎 成人, 高野 元, 原 良憲: "ネットワーク環境下における動画ハイパーメディア実装方式", in *Proceedings of Advanced Database System Symposium '94*, pp. 165-173, 1994 年 12 月.
- [小川 96] 小川 政行, 石川 佳治, 植村 俊亮: "圧縮映像データベースシステムにおける映像演算と実現手法", 情報処理学会第 106 回データベースシステム研究会 研究報告, 96-DBS-106, pp. 1-8, 1996 年 1 月.
- [増永 96a] 増永良文, 清光英成: "マルチメディアオブジェクト間の時間的関連記述のフレームワーク", 電子情報通信学会論文誌 D-II, Vol. J79-D-II, No. 4, pp. 492-501, 1996 年.
- [増永 96b] 増永良文, 野中 和明, 清光英成: "時区間論理に基づく時間的マルチメディアオブジェクトの同期再生スキームの実装", 電子情報通信学会論文誌 D-I, Vol. J79-D-I, No. 10, 1996 年, (掲載予定).
- [端山 96] 端山 貴也, 清木 康: "コンティニユアスメディアデータの表示・演奏を対象とした分散プロセス間同期機構と実現", 情報処理学会研究報告, 96-DBS-108, pp. 41-48, 情報処理学会, 1996 年 5 月.
- [八幡 95] 八幡 孝, 菅野 昭博, 下條 真司, 西尾 章治郎, 宮原 秀夫: "ニュース・オン・デマンドシステムにおけるニュースデータベース設計", 電子情報通信学会 信学技法, 第 95 巻, pp. 1-8, 1995 年.
- [福田 94] 福田 紀彦, 天笠 俊之, 金森 吉成, 増永 良文: "時区間に基づく拡張時間データモデル", in *Proceedings of Advanced Database System Symposium '94*, pp. 175-184, 1994 年 12 月.
- [有澤 93] 有澤 博, 由井 仁, 富井 尚志: "映像データベースシステムの構成の方式", in *Proceedings of Advanced Database System Symposium '93*, pp. 181-190, 1993 年 12 月.