

## 映像データベースシステムにおける映像演算の実装

7R-6

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

## 1. はじめに

本稿では、圧縮映像データベースシステムの実現を目的とした画像と音声操作のための意味構造(生, 場面, 合成)を定義し, 場面に對する演算, 合成に對する演算, 合成に用いる演算子を提案する. 本稿で画像とは動画像を指し, 画像と音声を合わせて映像と呼ぶことにする.

また, 画像と音声の同期を取るために空音声 (null audio) を導入し, 1プロセッサによる同期のずれを最小限に抑える手法を提案する.

これまでに開発したシステムでは, 画像に對して場面の索引付けを行い, データベース管理システムに格納しているの, インタフェースを介して場面画像だけを抽出することはできる. しかし, 音声は出力されない.

本システムでは, MPEG-1 で符号化された映像を対象とし, 画像と音声それぞれの索引を用いて場面を合成する演算子を実装する. これにより VOD として扱えるだけでなく, 所望の場面を合成するような新しい要求にも答えることができ, また音声に関する問合せにより画像だけでは困難な問合せ処理も可能となる.

実装は開発言語として C++ を, DBMS に UniSQL を, GUI には Tcl/Tk を用いる.

## 2. 意味構造

画像は意味のある構造として, 生画像, 場面画像, 合成画像から構成されるとし, 次のような特徴を持つ<sup>1), 2)</sup>.

**生画像:** 一つの線形時間軸を形成する画像素材 (野球中継そのもの).

**場面画像:** 生画像の任意の時間軸に利用者が意味付けを行って定義する画像 (ホームランの場面).

**合成画像:** 場面画像や合成画像を合成し, 利用者が意味付けを行って定義する画像 (ホームラン集, 珍プレー好プレー集).

同様に, 音声も意味構造として, 生音声, 場面音声, 合成音声から構成されるとし, 次のような特徴を持つ.

**生音声:** 一つの線形時間軸を形成する音声素材 (野球放送における音声そのもの).

**場面音声:** 生音声の中特定の音声に對して, 利用者が意味付けを行って定義する音声 (解説者の音声, 実況の音声).

**合成音声:** 場面音声や合成音声を合成し, 意味付けを行って定義する音声 (今日の野球の音声, インタビューにおける会話のみの音声).

## 3. 映像演算

## 3.1 演算の定義

場面に對する演算の定義を以下のように行う.

**場面属性定義演算 (define):** 場面における意味情報の定義.

**場面生成演算 (create):** 任意の時区間の意味付けによる場面の生成.

**場面削除演算 (delete):** 場面削除.

**場面更新演算 (update):** 場面情報の更新.

合成に對する演算の定義を以下のように行う.

**合成生成演算 (compose):** 場面もしくは合成の集合で構成.

**合成削除演算 (delete):** 合成削除.

**合成更新演算 (update):** 合成情報の更新.

合成に用いる演算子の定義を以下のように行う<sup>3), 4)</sup>.

**論理積 (intersection):** 時間的共通場面の抽出.

**論理和 (union):** 時間的共通場面を持つ二つの場面を, 共通場面の繰り返しを行わずに一連の場面として抽出.

**論理差 (difference):** 時間的共通場面を持つ二つの場面の差分場面を抽出.

**連結 (concatenation):** 任意の場面の連結.

## 3.2 画像と音声の同期

音声を 8bit, 8kHz でサンプリングする時, 音声の最小単位は 1/8000 s であり, 画像は 1 フレームが最小単位で 1/30 s である. このために音声に関する問合せを行って, 音声場面を取り出すとそれに対応する画像はフレームの途中から始まりフレームの途中で終わることになる. そこで, 画像と音声の同期を取るために, 空音声 (null audio) を導入する<sup>5)</sup>. これは, 開始時刻と終了時刻の情報のみを持つ空時区間オブジェクトである. これで画像と音声のずれている部分を埋めることで, 音声と画像の再生ストリームの開始時刻と終了時刻を一致させることができ, 同期再生が可能となる (図 1).

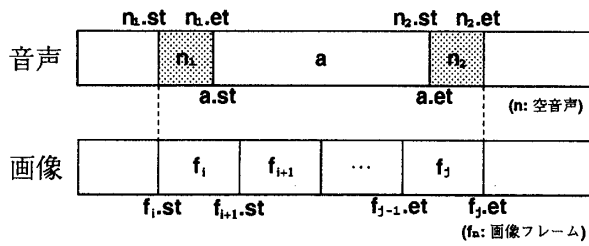


図1 画像と音声の同期

ここで、再生を以下のように行う。

- (1) 問合せにより音声  $a$  とそれに対応する画像フレーム  $f$  を抽出する。
- (2)  $f_i.st \leq a.st < f_{i+1}.st, f_{j-1}.st \leq a.et < f_j.et$  なるフレーム  $f_i, f_j$  を見つける。
- (3)  $n_1.st = f_i.st, n_1.et = a.st, n_2.st = a.et, n_2.et = f_j.et$  なる空音声  $n_1, n_2$  を生成する。
- (4) 画像と音声の同期を取って再生を行う。

### 3.3 同期制御

1 プロセッサで画像と音声を同期させて出力しようとする時、どうしてもずれが生じてしまう。このずれを最小限に留めるため、共有領域 (shared memory) を用いた制御を行う (図2)。共有領域は一方の処理が完了状態に有ることを示すフラグを持ち、開始時刻をセットできる。

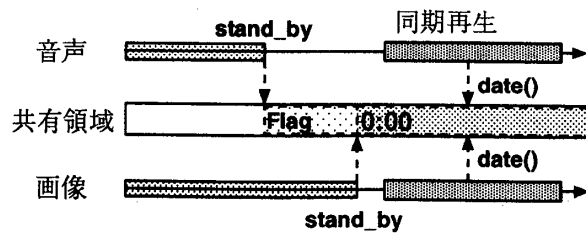


図2 同期制御

プロセッサは、画像に対して表示領域の確保を行い、音声に対してデバイスの I/O 制御を行う。それぞれの処理が完了次第、共有領域にアクセスし、待ち状態 (stand.by) であることを示すフラグを立て、もう一方の処理完了を待つ。待ち状態の間は、共有領域にアクセスし続け再生開始時刻の書き込みが無いかを見る。共有領域にアクセスした時に、もう一方が待ち状態であれば、再生開始時刻を書き込む。この時刻を元に画像と音声の同期再生を行う。これでも、プロセスの切替のためにはずれが生じるが誤差の範囲である。再生

開始後は一定間隔で共有領域にアクセスし、開始時刻と現在の時刻の差分を計算することでそれぞれの再生すべきところを確認することで同期を取る。

もし、両方が同時に共有領域にアクセスした時、排他制御 (semaphore) により優先度の高い方が先に処理をし、低い方はその処理待ち状態に入る。

### 4. 問合せ

問合せ演算としては、時間関連を用いた問合せ (query)、内容に基づく問合せ (content) を定義する。

これらの演算により、VOD による演算だけではなく、所望の場面映像を合成するような要求にも答えることができる。例えば、落合のホームランを打った場面を合成して、落合のホームランライブラリを作成できる。

また、音声に関する問合せにより場面を取り出すこともできる。例えば、インタビューの場面から会話部分のみの場面を取り出すことができる。

### 5. まとめと今後の課題

本稿では、画像と音声の構造化を行い、それらに対する合成のための演算、同期再生手法ならびに制御法を提案した。今後は、同期再生手法ならびに制御法の洗練を行い、講義映像に対して実装を試みる。

謝辞: 日頃から有益な御指導・御討論を頂く、植村研究室の皆様へ感謝致します。

### 参考文献

- 1) 小川 政行, 石川 佳治, 植村 俊亮: “映像データベースにおける代数的データモデルの提案”, 情報処理学会第 51 回全国大会, 5D-6, 1995 年 9 月
- 2) 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, 1996 年 1 月
- 3) Ron Weiss, et al.: “Content-Based Access to Algebraic Video”, *Proc. of the Intr. Conf. on Multimedia Computing and Systems*, May 1994, pp140-151
- 4) 堀内 優希, 友田 政明, 石川 佳治, 植村 俊亮: “映像データベースのための論理データモデルとその実装”, 電子情報通信学会第 6 回データ工学ワークショップ, 1995 年 3 月
- 5) 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一フレームワーク”, 電子情報通信学会論文誌, Vol.J79-D-II, 1996 年 4 月