

## 映像データベースのための同種メディアの統合

田中秀明 吉山雅彦 植村俊亮

奈良先端科学技術大学院大学 情報科学研究科  
〒630-01 奈良県生駒市高山町 8916-5

{hideak-t, masah-yo, uemura}@is.aist-nara.ac.jp

ネットワーク技術の発展と計算機の高速化によって、マルチメディア情報の利用が盛んになり、データベースで扱うことが可能となっている。TV、映画などで制作される映像は、膨大な数に達している。こうした環境の中で映像を扱うことのできるデータベースの研究がなされてきた。しかし、これまでの研究では同時間軸に対する複数映像を取り扱うことがなされておらず従来の索引付けの手法では、このような同時間軸に対する複数映像にまたがる問合せに対して有効な検索結果が得られない。そこで本稿では、この映像データベースに対し、同時間軸に対する複数映像を取り扱う拡張を行なって実装することを目的とする。その時の問題点や検索法について提案し、その構築方法について考察する。

## Integration of Movies with Same Time-axis for Video Database Systems

Hideaki TANAKA, Masahiko YOSHIYAMA, and Shunsuke UEMURA

Graduate School of Information Science  
Nara Institute of Science and Technology  
8916-5, Takayama, Ikoma, Nara 630-01, JAPAN

{hideak-t, masah-yo, uemura}@is.aist-nara.ac.jp

Recent progress in computer and network technology is making so called multi-media databases a reality. However, video database systems studied so far only dealt with information on single data stream. This paper discusses on how to manage video data on multiple data stream, proposes a method to find out relationships on different data stream, and discusses on implementation issues.

## 1. はじめに

映画、テレビなどの映像が1台だけのカメラで撮影されることは少ない。将来、個人でも、複数のカメラを用いて撮影し、それらの映像を組み合わせるようになるであろう。また、最近では映画やドラマの脚本の中には、ザッピングという手法を用いて同一時刻の複数の場面を見せることによって表現されることもある。しかし、これまでの映像データベース[2]では、一つのストリームについてのみに考えられており、同時刻に撮影された複数のストリームを格納、操作する場合について考えられていなかった。また、現在研究されているショットでの切分け[6]や、人物や対象物の動きから検出される索引付け[4]では、最終的に1巻に編集された映像を対象としているので、映像に映っているものだけしか検索することができない。つまり、それと同時に進行している現象についての関連性が保証されることはない。この関連性を記述することにより、これらの映像を検索できるはずである。

本研究では、同時刻に撮影された複数の映像のストリームのことを同時間軸映像ということにする。なおストリームとは、テープメディアにおいては1本のテープを、デジタル映像では、1ファイルことをいう。

## 2. 目標

ここでは、衛星会議を考える。衛星会議とは、衛星を用いて2箇所の地点からその基地局の様子の動画と音声を送り、会議に参加する数箇所の基地局が受信して会議を行なうシステムのことである。ただし、映像を送信している2箇所の基地局どうしのみ、会話を行なうことが可能となる。

会議参加者は、図1のように同時刻の瞬間を示した映像1と2を視聴することができる。図からも判



図1 アプリケーション例

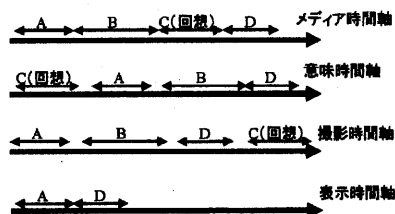


図2 メディア、意味、撮影、表示時間軸

るように、映像1と2は別々の会場を映し出しているため、その両方の映像共に登場している人物はいない。このため、従来の考え方では、二つの映像は何の関連性もないものになってしまう。

本研究では、このような映像1と映像2の二つの映像に関連性を持たせ、次の質問ができることを目標としている。

- i) t氏(図1:映像1左)とu氏(図1:映像2左から2番目)が会話している場面を検索せよ。
- ii) t氏が会話している人物が映し出されている場面を検索せよ。
- iii) t氏が映っている場面と同時刻の他の人の様子が映し出されている場面を検索せよ。

## 3. 映像における時間と時間関連

### 3.1 時間の分類

ここでデータベースの構築、検索の際に必要な映像の時間について考える。池谷らは、映像の時間をメディア時間、物語時間、撮影時間、表示時間の四つに分類した[7]。物語時間は、映画・ドラマを対象としている。そこで、あらゆる映像に対応するようにこれを意味時間として再定義する。意味時間を含めたこの四つの時間について説明する。

**メディア時間** 物理構造に対応していて、ビデオ映像など格納順序に沿った時間である。ストリームごとに、0秒から始まり、その映像長さ分の時間までが存在する。

**意味時間** 撮影者および映画等の制作者が意図する仮想の時間である。フィクションである多くの映画、ドラマは複雑な意味時間を持っていることが多く、映画等の時代設定、話のながれなどによって不連続かつダイナミックに変化する可

能性がある。そして、この意味時間は、何らかの情報の、すなわち場所や事象などで結びつけられる必要がある。

**撮影時間** その映像を撮影した時の時間である。例えば、1998年1月19日10時35分26秒のように撮影された実時刻である。

**表示時間** データベースで検索したシーンをどのような順序で表示するかを決定するために用意された時間である。

このように単に「時間」といっても4種類の時間が存在し、ビデオ映像はこれらの時間をもとにした時間軸がある。それぞれの時間についてももう少し具体的な例を図2に示す。

### 3.2 同時間軸映像における時間軸

ここまでは、1ストリーム上で考えきたが、以下では本研究の主題である同時間軸映像に対してそれぞれの時間を考える。例として、3本のストリームのうち2本が同時間軸映像である場合(図3)は、次のようになる。

**メディア時間** 3本のストリームがあるので、それぞれメディア時間は0秒から始まってそれぞれの映像の長さで終る。3本のストリーム中の時間をそれぞれ区別するためには、識別する機能が必要となる。

**意味時間** ストリームA、Bが同時間軸映像であるから、意味時間軸上でストリームA、Bが重なっている部分がある。ストリームA、Bの中の時間を区別するためには、識別する機能が必要となる。

**撮影時間** ストリームA、Bは同時に撮影されたものなので、撮影時間軸上でもストリームA、Bが重なっている。また、ストリームCがストリームA、Bより前に撮影されたものであると、このようになる。

**表示時間** 利用者が指定する表示方法によって表示時間の表し方が変わるので、同時に再生する場合と逐次に再生する場合が考えられる。

### 3.3 映像の分類

さらに我々は、メディア時間、意味時間、撮影時間の三つの時間から、次のように映像を分類する(図4)。

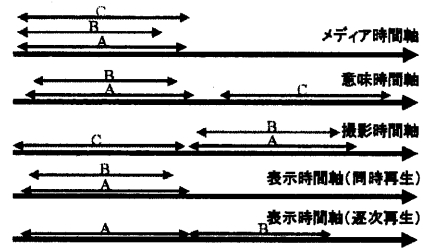


図3 同時間軸映像のメディア、意味、撮影、表示時間軸

映像タイプ	時間		
	メディア	意味	撮影
撮影素材	すべて一致		
中継	独立	一致	
ドラマ・映画	独立	独立	独立

図4 映像タイプと時間の相関

**撮影素材タイプ** カメラで撮影されたままで、編集を施していない映像があてはまる。つまり、一つのシーンだけで構成された映像がこのタイプに当たる。メディア時間、撮影時間、意味時間が一致している。

**中継タイプ** 撮影素材タイプの映像を複数組み合わせ、色々なカットやシーンを創り出した映像があてはまる。撮影時間の時間軸の切れ目は存在するが、メディア時間に対する撮影時間の入れ替わりは存在しないので、その時間軸は一方向で表せる。また、意味時間は撮影時間と一致する。

**ドラマ・映画タイプ** 中継タイプの映像よりさらに複雑化し、メディア時間、意味時間、撮影時間のそれぞれは、独立している。映画やドラマなどの映像は、ほとんどこのタイプに分類される。

## 4. 衛星会議の映像

本研究で扱う映像「衛星会議の映像」は、前節で述べた中継タイプに分類される。このことは、これから検索手法について考える上で、時間について複雑な場合を考えることがなくなる。さらに、この中継タイプを基礎にして、ドラマ・映画タイプがその発展形になると思われる。この映像の特徴として、す

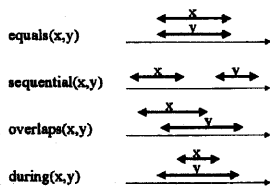


図 5 時間的関連述語

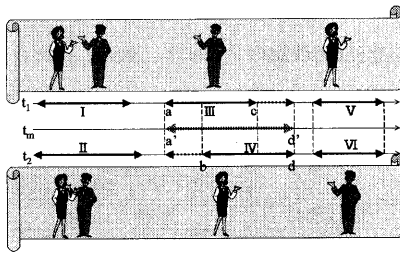


図 6 時区間の検索と演算

すべての映像において、厳密には異なっているが、同一時間にはほぼ同じと想定できる音声が含まれている。

### 5. 論理時間

検索の対象となる論理時間を定義する。論理時間とは、データベースの中で、一意に決まる時間である。データベースの制作者が、映像の意味時間に基づいて定義する時間となる。この論理時間を介して、複数の映像間の検索を可能にする。そのため基本的に論理時間は、先に述べた意味時間と一致する。したがって、撮影素材タイプの論理時間は、メディア時間、撮影時間とも一致することになり、中継タイプの論理時間は、撮影時間と一致する。

### 6. 時間的関連述語

時間と映像の関連を表し、演算を行なうために時間的関連述語 [1][2] を用いる (図 5)。時間的関連述語を用いて、映像の関係を明白にすることができる。

### 7. 時区間の検索と演算の問題

図 6 は、ある映像の例である。図上の  $t_1, t_2$  は、映像 1、2 それぞれのメディア時間軸を表している。そして、図上の  $t_m$  は、映像 1、2 に対する意味時間軸であり、映像 1、2 は、A さん (男性) と B さん (女性) が会話していることで同じ意味時間軸で表される。同じ映像の中で A さんと B さんが会話している

場合 (図 6: 左) は、たとえ意味時間が一致していなくても、二人が会話している場面を検索すると、場面 I、場面 II を結果として返すことができる。

次に、別々の映像の中で A さんと B さんが会話している場合 (図 6: 中、右) を考えてみる。前述の *equals* で表される状態 (図 6: 右) の時は、映像 1、2 の論理時間がどちらも同じなので問題はない。しかし、前述の *overlaps* で表される状態 (図 6: 中) の時は、検索されて表示される場面はさらに複雑になる。検索結果としては、場面 III ( $t_1: a \rightarrow c$ ) と場面 IV ( $t_1: b \rightarrow d$ ) となるが、論理時間軸  $t_m$  上では  $t_m: a' \rightarrow d'$  となる。しかし、このとき映像 1 のメディア時間軸  $t_1$  上の  $t_1: c \rightarrow d'$  は、「A さんと B さんが会話している場面」ではないという矛盾が起きる可能性がある。しかし、そうだからといってその共有 (AND) の部分の  $t_m: b' \rightarrow c'$  のみを取り出すのは、場面構成の意味がわからなくなる危険性がある。つまり、厳密な意味で「A さんと B さんが会話している場面」というを検索するのではなく、「A さんと B さんが会話しているであろう場面」が検索される方が現実味があると思われる [3]。

## 8. 実装

### 8.1 映像データベースに用いる表

OID	登場人物	ST-No	開始時間	終了時間
0001	植村	03	07299	08345
0002	田中	01	07123	08233
0003	吉山	03	12383	13943
....		.....		
0193	田中	05	19123	34248

ST-No: ストリーム番号

図 7 表の例

映像データベースの中で管理する表を図 7 のように時間の関係だけで作成する。この表では、登場する人物について着目している。その人物が映像に登場する時区間を「ST-No」(ストリーム番号: ストリームを識別するための ID)、「開始時間」、「終了時間」で表している。「開始時間」、「終了時間」は、論理時間を用いて格納されている。

### 8.2 SQL による問合せ

Select 映像 from 表 where A talk\_with B

この例は、A と B が話している映像を検索するための問合せである。*talk\_with* の部分を変えると色々な

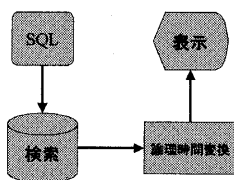


図 8 検索の流れ

関連について対応することができる。また、**映像**の部分、時間的関連述語によって求められる「ST-No」、「開始時間」、「終了時間」の組が検索されることになる。

### 8.3 検索の流れ

検索の流れを、図8に示す。まず、利用者はSQLによって検索したい映像を表現する。SQLは、すべての利用者が使いこなすことは難しく、実際にはGUIを用いた検索インタフェースにすることが望ましい。そして、SQLで書かれた問合せから表を後で述べるように検索し、それに当てはまる「ST-No」、「開始時間」、「終了時間」を計算する。ここで求められた「開始時間」、「終了時間」は、論理時間であるから、これをストリーム毎に持っているメディア時間への変換が必要となる。映像が中継タイプの場合、この変換は、意味時間が撮影時間と変わらないため簡単になるが、映画・ドラマタイプの場合、意味時間が複雑になればなるほど時間変換は複雑になる。

こうして、それぞれのメディア時間で表された「開始時間」、「終了時間」を使って映像を再生することができる。

### 8.4 目標の問合せの検索法

ここまで述べたことを使って目標とした問合せの検索方法を考える。

i) t氏とu氏が会話している場面

- (1) 「t氏」と「u氏」の項目をそれぞれ表の「登場人物」から検索する。
- (2) それぞれの結果の論理時間の関係が *sequential* ではない関係のもの組を作る。
- (3) その関係の組に、問合せに合致する映像が含まれている。

ここで、検索されたものは、同じ時間を共有しているt氏とu氏であって、二人が会話してい

る場面であるかは、先に定義した表からでは検知することができない。

ii) t氏が会話している人物が映し出されている場面

- (1) 「t氏」の項目を表の「登場人物」から検索する。
- (2) その検索された論理時間の関係が *sequential* 以外の関係のものを計算して導き出す。
- (3) その導き出されたものの「登場人物」を基に再検索すると、その中に問合せに合致する映像が含まれている。

ここで、検索されたものは、t氏と同じ時間を共有している人物であって、検索された人物がt氏と会話している場面があるかは、先に定義した表からでは検知することができない。

iii) t氏が映っている場面と同時刻の他の人の様子が映し出されている場面

- (1) 「t氏」の項目を表の「登場人物」から検索する。
- (2) その論理時間が含まれる映像が問合せに合致する映像として検索される。

#### 8.4.1 表の考察

先に定義した表では、前節で示したように「会話」という動作を含めて検索しようとしても必要条件を満たしたものが検索されるだけで十分ではない。これが分かるためには、「会話」など動作を示した項目が必要となる。さらに、映像上に複数の人物がいる場合には誰と会話しているかという「動作の対象(相手)」も指定しないと会話していない人物とも会話しているように検索される可能性がある。

### 8.5 表の拡張

厳密な検索ができるように、先の表を図9のように拡張することを考える。先の表と同じようにその人物が映像に登場する時区間を「ST-No」、「開始時間」、「終了時間」で表し、さらにその時区間で関連を持つ相手とその行為を「相手」、「行為」で表す。この表の「開始時間」、「終了時間」も、論理時間を用いて格納されている。

#### 8.5.1 拡張した表での目標の問合せ検索法

拡張した表で目標とした問合せの検索方法を考える。

OID	登場人物	ST-No	開始時間	終了時間	行為	相手
0001	植村	03	07299	08345	talk	田中
0002	田中	01	07123	08233	talk	植村
0003	吉山	03	12383	13943	—	—
....			.....			
0193	田中	05	19123	34248	ans	吉山

ST-No: ストリーム番号

talk: 会話 ans: 応答

図 9 拡張した表

i) t 氏と u 氏が会話している場面

- (1) 「t 氏」と「u 氏」の項目をそれぞれ表の「登場人物」から検索する。
- (2) それぞれの結果の論理時間の関係が *sequential* ではない関係のもの組を作る。
- (3) その関係の組の中で「会話している」関係のものを「行為」と「相手」から計算して導き出す。
- (4) その結果、問合せに合致する組の映像が検索される。

ii) t 氏が会話している人物が映し出されている場面

- (1) 「t 氏」の項目を表の「登場人物」から検索する。
- (2) それに出てくる「相手」を用いて表の「登場人物」から検索する。
- (3) それから得られた論理時間から問合せに合致する映像が検索される。

iii) t 氏が映っている場面と同時刻の他の人の様子が映し出されている場面

検索方法は、拡張する前の表と同じようになる。

### 8.5.2 拡張した表についての考察

拡張する前の表 (図 7) に比べて拡張した表 (図 9) は、問合せに対してより正確に結果を示すことができるようになる。しかし、この拡張した表を作成すること自体が極めて手間のかかる困難な作業となることを考えるとあまり現実的ではない。

## 9. まとめと今後の課題

映像データベースに対し、同時刻軸に対する複数映像を取り扱う拡張を行なうと共にその検索法について提案し、その時に生ずる問題点やその構築方法

について考察した。提案した論理時間や SQL による問合せを整理するとともに、論理時間だけの表を作成し、オブジェクト指向リレーショナルデータベースを用いて実験システムを構築していく計画である。

### 参考文献

- [1] J. F. Allen. Maintaining Knowledge about Temporal Intervals. *Comm. of the ACM*, Vol. 26, No. 11, pp. 832-843, November 1983.
- [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, 96-DBS-106, pp. 1-8, January 1996.
- [3] 吉山雅彦, 田中秀明, 植村俊亮. 映像データベースのための異種メディア間の演算. 情報処理学会第 114 回データベースシステム研究会研究報告, January 1998.
- [4] 宮森恒, 粕谷英司, 富永英義. 動作語を用いた問合せによる映像検索方式. 電子情報通信学会論文誌, Vol. J80-D-II, No. 6, pp. 1590-1599, 1997.
- [5] 天笠俊之, 鈴木邦彦, 有次正義, 金森吉成. 時区概念モデルを実装した時区間クラス. In *Advanced Database Symposium'97*, pp. 59-66, December 1997.
- [6] 上田博唯, 宮武孝文, 炭野重雄, 長坂晃朗. 動画像解析に基づくビデオ構造の視覚化とその応用. 電子情報通信学会論文誌, Vol. J76-D-II, No. 8, pp. 1572-1580, 8 1993.
- [7] 池谷勇一, 飯田卓郎, 石丸知之, 中森 真理雄, 植村俊亮. 動画メディアのためのデータモデルの提案. 信学技報 DE93-53, 電子情報通信学会, 11 1993.