

同一時間軸上の複数メディアの演算

5G-3

吉山雅彦 田中秀明 植村俊亮

奈良先端科学技術大学院大学 情報科学研究科

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

1. はじめに

DVC,DVC-PRO,D-5など映像機材のデジタル化と、大容量HDDや高速ネットワークなどの計算機環境の飛躍的發展により、ノンリニア編集が可能になってきている。

またビデオデータをデジタル化してサーバに置き、要求されたデータをクライアントマシンに流し再生するVODも現実のものになりつつある。

しかしVODはビデオ映像の単純な連続の再生の要求であり、また用いられる連続メディアサーバ¹⁾もマルチメディアデータベース²⁾とは異質なものである。近年では連続メディアデータの中から意味のある部分をランダムに呼び出すデータベースの要望が高まりつつある。

我々はこのデータベースのことを映像データベースと呼び、これらの実装と検証を行ってきた。この結果から得られたデータにより、利用者の所望する検索結果を返すことのできる時区間演算について検討する。

2. 映像の内容スキーマ

映像の内容スキーマを考える際に、映像の時間軸をどのように扱えばよいかを分類する。

1. 討論会や会議等の場合は、スイッチング切替えなどの手法を用いることはあっても、映像内容の時間と撮影時の時間とが同一に流れている。
2. ニュースの場合、ニュースの話題それぞれについて同じ場所であれば時間軸は一方に流れており、また場所が異なる場合には並列化などが行われる。ニュースの話題同士に関してはほとんど相関関係がなく、別の内容として扱うことができる。
3. 旅行見聞録などの場合では、一元撮影の場合は場所の移動や時間の不連続などはあっても時間は一方に流れている。多元撮影の場合は複数の場所があるが、時間はそれぞれの撮影において一方

に流れており、複数の時間軸を用意することにより時間演算が可能になる。

4. 映画やドラマの場合では、内容は複雑になり複数の時間、空間がいろいろと混ざり合い、独自の時空間をなしている。次元ドアなどを通して現在・過去・未来と時空間をかけめぐるものさえある。

これらのことから、映画やドラマの場合では内容からの検索はかなり複雑なものとなり、すべての映像に対する包括的な内容スキーマの製作は困難である。しかしニュース、旅行記などは複数の時間軸における時間とそれぞれの時間軸に割り当てたパラメータを基準とした索引付けが可能である。これにより時間軸ともう一つのパラメータで時間軸を一つの場面として考えることができる。会議などに関しては元々の時間軸で場面を扱うことが可能である。

3. 検索用索引の最小単位

画像に対しては非常に正確に索引を付けることが可能である。例えばNTSC方式の映像は、1秒間に30フレームの静止画からなるため最高で30分の1秒を最小単位とした索引をつけることが可能である。また音声に関しては96KHzのDAT等を使うことにより、最高で約10万分の1秒単位の索引をつけることが可能である。

これらのことから画面と音声の同期をとるためには、検索結果として返す時間の最小単位は30分の1秒を考えれば充分である。

しかし本研究では記憶容量の制約から当分の間圧縮映像を対象とせざるを得ないと判断し、圧縮映像の乱呼び出しに取り組んできた。MPEG1で圧縮した映像における乱呼び出しの最小単位はGOPであり約2分の1秒になる。

4. プロトタイプシステムによる検証

時間の流れが一定であり内容スキーマの構築が比較的容易であると思われる多地点間遠隔会議等を収録した映像データベースのプロトタイプシステムを、データに時区間を用いた索引を付加して構築した。

Operations on the Time Axis of Multi Media Infomation
Masahiko YOSHIYAMA, Hideaki TANAKA, and Shun-
suke UEMURA
Graduate School of Information Science, Nara Institute of
Science and Technology (NAIST)

またこの度検索する索引の単位時間として索引付けする人間にもわかりやすい1秒を採用した。

あまりにも利用者の主観に比重がおかれている検索では、検索結果が利用者の所望の要求と異なる可能性が高いと考えられる。それ故にこれらよりも具体的な検索要求が望まれる。

このことを考慮して今回は画像での検索、音声での検索、内容での検索（キーワード検索）の三つの検索項目を実装しこれらの検索に関する実証実験を行った。

この実験より以下のことが確認された。

1. 画像での検索では「A氏の写っている場面」のように検索を行う。これによりA氏が画面に写っている場面が得られたが、喋っている人物がA氏と関係がない場合がある
2. 音声での検索においては「A氏の話している場面」のように検索を行う。これによりA氏が話している場面が得られるのだが、このとき画面上では全然A氏と関係のない画像である場合がある
3. 内容での検索（キーワード検索）では「ワークフローに関して議論している場面」のように検索を行う。今回は画面に写し出されているキーワードを元に索引を付けたが、これだけでは所望の映像が得られない場合がある

画像や音声に対しては要求に合致した検索結果を返すことができた。しかし画像と音声の検索では所望の映像を返せるが、内容による検索に関しては何をもって内容とするかが問題となることがわかった。

画面に写っているキーワードだけでは所望の映像を得ることはできない。また場面中の人物により内容を決めるという方法もあるが、そのキーワードの權威の人が話していればいいというでもない。

音声により内容を決定する場合でも、キーワードを話しているシーンだけやキーワードの前後のみを取り出せばよいのではないが、画面よりも音声によって内容を決める方がより所望の映像に近い検索結果を返すことがわかった。

5. 新たな演算子の定義

これらを元にして、意味のある検索を可能にするために画面と音声の両方を使った時区間演算子として新たにEXTRACTを定義する。

これはEXTRACT(Object, Video, Audio)と表記し、ANDやORとは異なる操作を行う演算子である。

1. 画像と音声の両方に対してObjectが含まれる時区間の組をandによって求める
2. 得られた時区間の組のそれぞれの時区間に対して、同じ属性(Video, Audio)で隣接する時区間も合わせて一つの時区間と置換する。

例えば下図の例では、ある映像オブジェクトを抜き出すという演算は、映像と音声の関係が下図のような関係にある時に、画像と音声の重なり(and)部分はc,eであるがc,d,eが音声で繋がっているので一連の内容とみなし、またbとc,eとfが画像で繋がっているのでb,c,d,e,fのすべてを一連の内容とみなす。しかしORでは含まれるはずのa,gについてはこの演算では含まれない。

これにより意味のある映像オブジェクトの抜き出しが可能となる。

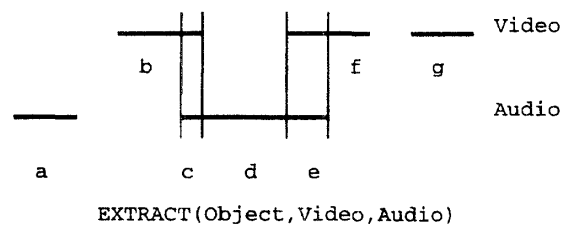


図1 映像オブジェクトの取り出し

6. まとめ

内容の決定のために複数の時間軸を考え、それぞれに対して独立した時間を設定し、それぞれの時間軸に対して場面パラメータを導入する。検索の際にはこの場面パラメータと時間とを合わせて時区間演算を行い、また今回導入したEXTRACT演算により内容を指定し、所望の映像を得ることができると考えられる。

参考文献

- 1) Banu Ozden, Rajeev Rastogi, Prashant Shenoy, AviSilberschats, *Fault-tolerant Architectures for Continuous Media Servers*, Proceedings SIGMOD'96, pp. 79 - 90, June 1996
- 2) 増永 良文, マルチメディアDMBSのためのデータモデルとシステムアーキテクチャ, 情報処理学会研究報告 データベースシステム 50-5, 1985, 50-5
- 3) 山本 憲男, 高倉 弘喜, 植村 俊亮, 映像データベースシステムにおける映像演算及び画像と音声の同期, 奈良先端科学技術大学院大学修士論文, March 1997