

同一時間軸上の複数動画像の検索法

5 G-4

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

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

1. 概要

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く、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。現在研究されているショットでの切分け⁴⁾や、人物の動きから検出される索引付けでは、最終的に一巻に編集された映像を対象としているので、動画像に映っているものに対して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。会議の映像の例では、ある人が話している場面とそれを聞いている人の場面がある。一つの映像の中に同時にこの場面が表現されていれば、従来の方法で関連性は生まれるのであるが、別々のカメラで撮られた映像であれば、関連性を検出することができない。それを改善できれば、新しい知識を見つけ出すことが可能となる。本稿では、この関連付けの方式を提案することを目的とする。

2. 実現できる効果

簡単に考えるため、図1のようにA, B, C, D, E, F, Gという七つの場面があって、a, b, c, dという4人の登場人物がいる場合を考える（図1）。2本の直線は動画像を表し、右方向に時間が流れていることを示している。これまでの索引付けでは、「人物cが登場している場面」を検索すると、D, Eが結果として返される。そして、「人物aが登場している場面」を検索すると、A, B, Fが結果として返される。ここで、aとcが会話していることが分かれば、動画像1と動画像2の関連性が生まれる。つまり、cが話している相手がaであるという情報が何らかの手段で付け加えられると、「cが話している相手が映っている場面」という検索の結果が、aが映っている場面A, B, Fとなり、さらに実際にcがaと話している場面は、a, cの映っているB, D, Eという結果を返すことが可能になる。さらに、aは、動画像1と2のど

ちらにも登場しているが、aの同時刻で異なる角度での様子を知ることができ、それを検索することも可能となる。

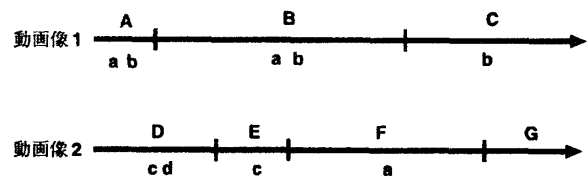


図1 イメージ図

3. 動画像における時間と時間関連

3.1 時間の分類

池谷らは、動画像の時間をメディア時間、物語時間、撮影時間の四つに分類した⁵⁾。本稿において特に必要であるメディア時間、物語時間、撮影時間について述べる（図2）。

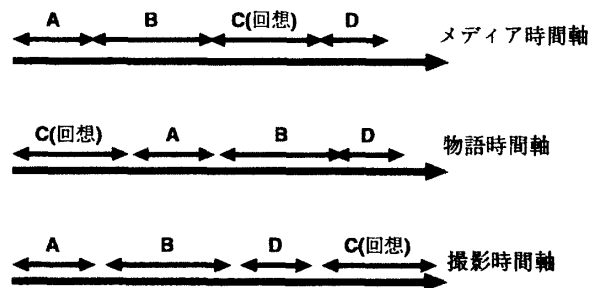


図2 メディア、物語、撮影時間軸

メディア時間 物理構造に対応していて、ビデオ映像の格納順序に沿った時間である。

物語時間 撮影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、ドラマはこの物語時間を持っている。

撮影時間 その映像を撮影した時の時間である。

3.2 動画像の分類

ここで我々は、メディア時間、物語時間、撮影時間の三つの時間から、次のように動画像を分類する（表1）。

The Method of Retrieval for Movies on the Same Temporal Axis
Hideaki TANAKA, Masahiko YOSHIYAMA, and Shunsuke UEMURA
Graduate School of Information Science, Nara Institute of Science and Technology(NAIST)

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動画像。メディア時間、撮影時間、物語時間が一致している。

中継タイプ 撮影素材タイプの動画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動画像。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、その時間軸は一方方向で表せる。また、物語時間は撮影時間と一致する。

ドラマ・映画タイプ 中継タイプの動画像よりさらに複雑化し、メディア時間、物語時間、撮影時間のそれぞれは、独立している。商業ベースの動画像は、ほとんどこのタイプに分類される。

4. 本手法実現のための方法

本研究で扱う動画像は、前節で述べた中継タイプに分類される「衛星会議の映像」である。この動画像の特徴として、全ての動画像において、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。

4.1 仮想時間

検索に使われるキーとなる仮想時間を定義する。仮想時間とは、データベースの中で、一意に決まる時間であり、この仮想時間を介して、複数の動画像間の検索を可能にする。基本的に仮想時間は、先に述べた物語時間と一致する。したがって、撮影素材タイプの仮想時間は、メディア時間、撮影時間とも一致することになり、中継タイプの仮想時間は、撮影時間と一致する。

4.2 動画像の同一時間軸への射影

複数動画像同志の時間軸の同期をとるために、音声によるDPマッチング³⁾を用いる。これにより仮想時間を一意に決めることが可能となる。

動画像タイプ	時間			複数動画像の取扱い
	メディア	物語	撮影	
撮影素材	すべて一致			なし
中継	独立	一致		あり
ドラマ・映画	独立	独立	独立	あり

表1 動画像タイプと時間の相関

4.3 時間関連述語

時間と動画像の関連を表し、演算を行なうために時間的関連述語¹⁾²⁾を用いる。時間関連述語には、動画像の時間的一致関連を表す *equal*、動画像の時間的不一致関連を表す *sequential*、動画像の時間的重複関連を表す *overlap*、動画像の時間的包含関連を表す *during* がある。時間的関連述語を用いることにより、動画像の関係が明白となり複雑な事象にも対応することができる。

4.4 SQL への対応

```
Select 映像 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関連について対応することができる。また、`映像` の部分は、時間的関連述語によって求められる仮想時間が検索されることになる。

5. まとめと今後の課題

これら仮想時間や時間的関連述語を整理し、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. 映像データベースのための演算体系と圧縮情報構造に関する研究. Master's thesis, 奈良先端科学技術大学院大学 情報科学研究科 情報システム学専攻, February 1996.
- 3) 古井貞熙. デジタル音声処理. 東海大学出版会, 1985.
- 4) 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.
- 5) 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.